GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 00:47:10 ; Search time 2103.99 Seconds
           (without alignments)
           15028.587 Million cell updates/sec

Title: US-09-786-135-4
Perfect score: 1511
Sequence: 1 tccagagtgctgaatttctg.........atcctatgcaataaagtgca 1511

ng table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

searched: 1797656 seqs, 1046326829 3 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
        1: gb_ba:*
        2: gb_htg:*
        3: gb_in:*
        4: gb_om:*
        5: gb_ov:*
        6: gb_pat:*
        7: gb_ph:*
        8: gb_pl:*
        9: gb_pr:*
       10: gb_ro:*
       11: gb_sts:*
       12: gb_sy:*
       13: gb_un:*
       14: gb_vi:*
       15: em_ba:*
       16: em_fun:*
       17: em_hum:*
       18: em_in:*
       19: em_mu:*
       20: em_om:*
       21: em_or:*
       22: em_ov:*
       23: em_pat:*
       24: em_ph:*
       25: em_pl:*
       26: em_ro:*
       27: em_sts:*
       28: em_un:*
       29: em_vi:*
       30: em_htg_hum:*
       31: em_htg_inv:*
       32: em_htg_other:*
       33: em_htgo_inv:*

(handwritten: Cite hits 1-97?)

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB	ID	Description
1	1294.4	85.7	1903	9	AB051125	AB051125 Macaca fa
2	1268.8	84.0	3319	2	BC001200	BC001200 Homo sapi
3	1217.8	80.6	151032	2	AC061705	AC061705 Homo sapi
4	1217.8	80.6	229586	2	AC048331	AC048331 Homo sapi
5	1214.6	80.2	152623	2	AC024530	AC024530 Homo sapi
6	1211.4	80.2	298408	2	AC092931	AC092931 Homo sapi
7	1110.4	73.5	2814	9	AB047829	AB047829 Macaca fa
8	471.8	31.2	198696	2	AC087898	AC087898 Mus muscu
9	471.8	31.2	199359	2	AC074027	AC074027 Mus muscu
10	471.8	31.2	219564	2	AC073708	AC073708 Mus muscu
11	271.8	18.0	173803	9	CNS01DV2	AL133448 Human chr
12	271.8	18.0	208953	9	CNS01RH3	AL160314 Human chr
13	103.4	6.8	80767	9	HS745E8	AL020998 Homo sapi
14	103.4	6.8	146722	9	AL139254	AL139254 Homo sapi
15	100.4	6.6	2116	9	AK057346	AK057346 Homo sapi
16	95	6.3	61503	2	AC107845	AC107845 Mus muscu
17	84.8	5.6	125020	9	AF429315	AF429315 Homo sapi
18	70.8	4.7	125020	9	AF429315	AF429315 Homo sapi
19	65.6	4.3	168554	2	AC096863	AC096863 Mus muscu
20	63.8	4.2	207064	2	AC106680	AC106680 Rattus no
21	62.6	4.1	197669	2	AC084064	AC084064 Homo sapi
22	61.2	4.1	176370	2	AC105829	AC105829 Rattus no
23	60.8	4.0	10732	6	E32986	E32986 Gene encod
24	59	3.9	7218	6	I66494	I66494 Sequence 14
25	57.6	3.8	205914	6	AC027682	AC027682 Homo sapi
26	57	3.8	55021	2	AC107821	AC107821 Mus muscu
27	57	3.8	55536	2	AC090310	AC090310 Homo sapi
28	54.8	3.6	129600	2	AC097133	AC097133 Rattus no
29	54.8	3.6	219952	2	AC084804	AC084804 Mus muscu
30	54.4	3.6	176877	2	AP003902	AP003902 Oryza sat
31	54.4	3.6	298166	2	AC087563	AC087563 Homo sapi
32	54.4	3.6	300695	2	AC079431	AC079431 Mus muscu
33	54.2	3.6	9389	1	STE250581	AJ250581 Streptomy
34	54.2	3.6	166341	2	AC013381	AC013381 Homo sapi
35	54	3.6	4233	2	HSU75308	U75308 Human TBP-a
36	54	3.6	90935	2	AP003917	AP003917 Oryza sat
37	54	3.6	161307	2	AC084058	AC084058 Homo sapi
38	54	3.6	176986	2	AC108671	AC108671 Homo sapi
39	53.8	3.6	92310	2	AC108380	AC108380 Pan trogl
40	53.8	3.5	94914	2	AC105592	AC105592 Rattus no
41	53.2	3.5	303091	2	AC084799	AC084799 Mus muscu
42	52.8	3.5	104341	2	AC099456	AC099456 Rattus no
43	52.6	3.5	82815	2	AC062001	AC062001 Homo sapi
44	52.2	3.5	85434	2	AC066610	AC066610 Homo sapi
45	52.2	3.5	221852	2	AC069214	AC069214 Homo sapi

ALIGNMENTS

RESULT   1
AB051125

```
LOCUS       AB051125                1903 bp    mRNA    linear   PRI 11-OCT-2001
DEFINITION  Macaca fascicularis brain cDNA, clone:QccE-20281.
ACCESSION   AB051125
VERSION     AB051125.1  GI:11231112
KEYWORDS    fis (full insert sequence).
SOURCE      Macaca fascicularis adult male cerebellum cortex cDNA to mRNA,
            clone_lib:macaque brain cDNA library QccE clone:QccE-20281.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1903)
  AUTHORS   Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirata,M.,
            Suto,Y., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
  TITLE     Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
            chromosomes
  JOURNAL   Gene 275 (1), 31-37 (2001)
  MEDLINE   21458551
REFERENCE   2  (bases 1 to 1903)
  AUTHORS   Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
  TITLE     Direct Submission
```

```
JOURNAL    Submitted (14-NOV-2000) Katsuyuki Hashimoto, National Institute of
           Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
           1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
           (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,
           Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
COMMENT    Lab host:  TOP10
           Vector:    pME18S-FL3 (Acc.NO. AB009864)
           R. Site1:  DraIII (CACTGTGTG)
           R. Site2:  DraIII (CACCATGTG)
           Description: 1st strand cDNA was primed with an oligo(dT) primer
           [ATGTGGCGGCTTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
           using specific 5' and 3' primers and amplified by PCR. The PCR
           product was digested with SfiI and size selection was performed to
           exclude fragments <1.5kb.The SfiI-digested PCR product was cloned
           into distinct DraIII sites of pME18S-FL3. XhoI just outside
           the DraIII sites can be used to isolate the cDNA insert. Libraries
           were constructed by Sugano et al.(University of Tokyo, Institute of
           Medical Science). Custom primer used for sequencing
           ( 5' end primer [CGACCTGCAGCTCGAGCACA] ).
           3' end primer [CGACCTGCAGCTCGAGCACA] ).
FEATURES             Location/Qualifiers
     source          1..1903
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
                     /clone="QccE-20281"
                     /sex="male"
                     /tissue_type="cerebellum cortex"
                     /clone_lib="macaque brain cDNA library QccE"
                     /dev_stage="adult"
     CDS             809..1591
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAB18151.1"
                     /db_xref="GI:11231113"
                     /translation="MIPTAGLQGGSPAGAWDSDRNGNSKCALGDAATPMEGPRCPPPR
                     SPSRLSLGRRHKJCRPDLGQANNSEGIDHDYLPLVRLQEAPGSFRLDAPFCAAVRISQ
                     ERLCRASPFAVHRASLSPTSASLPWALLGPVGVGQDSATAFCSPSPSSGSEPGPGVDS
                     GRWGSDTEASEAGEGLGGTDLRGRTWATAVALAWLEHRCAAAFGEWELTAAKADCWLRA
                     QHLPDGDLAALKAAARGLFLLLRHWDQNLQLHLLCYSPANV"
BASE COUNT      374 a    597 c    537 g    395 t
ORIGIN

Query Match             85.7%;  Score 1294.4;  DB 9;  Length 1903;
  Best Local Similarity 91.1%;
  Matches 1436;  Conservative 0;  Mismatches 71;  Indels 69;  Gaps 3;

Qy      1 tccagagtgctgaatttctgggcagccaggggggctctgctctgctatgggttgaagatt 60
          ||||||||||||| ||||||||||||||||||||| |||||  ||||| |||||||  ||
Db    299 TCCAGAGTGCTGAA-TTCTGGGCAGCCAGGGGGGCCCCCTTGTTGTCTGCTACGGGTTGAAGGTT 357

Qy     61 cagctgatcacttcttgtcaagttggggagtctgagctggttctgggcctagcaaggc 120
          |||| |||| ||||||||||||| || ||||||| |||| || ||||||||||||||||
Db    358 CAGCCTGACCACTTCTTGTCAAAGTCGGGGAGTCGGGAGTCGAGCTGGTTCTGGGCCTAGCAAGGC 417

Qy    121 aattactctgttaaagtcgatagagggagaagctgggagagaacactgcaatttttccta 180
          |||| |||||||||||||||||||||||||||||||| |||| ||| ||||||||| |||
Db    418 AATCTACTCTGTTAAAGTCGATAGAGGGAGAAGCTGGGAGAATACTGCAATTTTTTCTA 477

Qy    181 tgagccctctgtagagggtgggtgtggggggccacagactgtctccttggttccacagg 240
          |||||||| ||||||| |||| |||| |  |||||||||||||||||||||||||||||
Db    478 TGAGCCCTCTGTAGAGGGTGGGTGTGCTGAGAGGGTGTAAGGGTTCCACAGACCGGTCCTTGGTTCCACAGG 537

Qy    241 ccatgccgagtagtgctggcttcgagccttcagacaagtaaggtcagctctgcccctc 300
          |||||||| |||| ||||||||| || |||||||||||||||||||||||||||||| |
Db    538 CCATGCCGCCCGAGGTGCTGCCTGCCTTCCAGCCCTTCAGACACAGTAAGGTCAGCTCTGCCCCTTC 597

Qy    301 ctgcttcacttgcctgtagctgcgtgatgctactactaggaggtcctgctggggccct 360
          |||||||||||||| ||||||||||| |||| |||||| |||||||||||||||| |||
Db    598 CTGCTTCACTTGCCCGTAGCTGCGTGCTGCGTGGATGCTACCACTAGGAGGTCCTGCTGGGGGTCCT 657

Qy    361 gcaggtgcagcctcagagcccgctgagccccaggaacccctcctgctctcacagcca 420
          ||||||| |||||||||||||||||||||||| ||||||||||| |||||||||| ||
Db    658 GCAGGTGTGCAGCTCAGAGCCAGCTGAGCCCCGCTCAGCCCCCAGGAACCCCTCCTGCCGCTCACAGCGCCG 717

Qy    421 tctagatgcagctcctctgccccactgttgtctactctaa-------- 459
          |||||||||||||||||||||| |||||||||||||||||
Db    718 TCTAGATGCAGCTCCTCTGCCCTCCACTGTTGTCTACTCTAAAAGGTAGGAAAAGGTAGGGGC 777

Qy    460 --------------------------------------------aggacttcagagagg 474
                                                      ||||||||||||
Db    778 ACTTGGACTTAGAGACCACCCGCTGGCACTGATGATCCCACTGCAGGACTTCAGGGAGG 837

Qy    475 ctctccagcagcagccggactggaccagacaaaatgcaacccaagcgtgctttggggga 534
          |||||||||||||||||||||||||||||||||| ||||||||| |||| |||||| ||
Db    838 CTCTCCAGCAGCAGCCGGACTGGACCAGACAGAAATGCAACTCCAAGTGTGCTTTGGGGA 897

Qy    535 ccctgccactccacggaaggtctcgccgccacctcccgtcctccctgtcggctcag 594
          | ||||||||||| ||||||||| |||||||||||||||||||||  |||||| |||
Db    898 CGCTGCCACTCCCATGCAGGTCCTGCGCCACCTCCCGTTCCCCGTCTCCCCTCTCGGCTTAG 957

Qy    595 catgggcggcgtcacaaactctgtagccctgaccagcggcaggcaacaacagtgaagg 654
          | |||||||| ||||||||||| ||||||||||| ||| |||||| |||||||||||||
Db    958 CCTGGGCGGCCGTCACAAACTCGTAGACCCTGACCTGGCCCAGGCCAACAACAGTGAAGG 1017

Qy    655 cagcgaccatgactacctgccctggtcggctcaggagaggcaccaggctccttccgcct 714
          || |||||||||||||||||||||| ||||||||||||||| |||||||||||||||||
Db   1018 CATCGACCACCATGACTACCTGCCCTGGTGCGGCTCAGGAGAGGCCCCAGGCTCCTTCCGCCT 1077

Qy    715 ggacggcgccttctgcgcgctgtgcgatctgcaggagcgcctctgcgcctgcc 774
          ||||||||||| ||||||| |||||| ||||||||||||||||||||||  |||
Db   1078 GGACGGCGCCCTTCTGCGCGTGTGCGCATCTGCAGGAGCGCCTCTGCGCGCCTCGCC 1137

Qy    775 ctttgcgtgcaccgcgccagcctcagcccacctcggcctcattgccctggcacttct 834
          ||||||||||||| |  | |||||||||  |  |||||||||||||||||||| |||||
Db   1138 CTTTGCCGTGCACCGTGCACCGTGCCAGCTCGGCCTCATTGCCCTGGCCACTTCT 1197

Qy    835 gggccctggttggtggccagggtgacagtggacagtggccaggtggcgggtccccagtc 894
          || ||||||| |||||  |||||||||||| |  | |  |  ||  |||| |||||||||
Db   1198 AGGCCCCTGGTGTTGGTTCAGGGTGACAGTGCCACAGCCTTTCTGCAGCCCGTCCCCCAGCTC 1257

Qy    895 gggctctgaggggccaggcagtggcgggtctcagacaccgaggcctcga 954
          || ||||| ||||| ||| |||||  ||| ||||| ||||||||||||||
Db   1258 GGGCTCCGAGGGTCCGGACAGTGGACAGTGGACAGGTGGGGGGCTCAGATACCGAGGCCTCGGA 1317

Qy    955 tgggccgaaggctgggcgcgaccgacctgggggccggaccgtggccactgccgtagc 1014
          |||||||||||||||||||| |||||||| |||| ||||||| ||||||||||| |||
Db   1318 GGGGGCGGAAGGGCTGGGCGCGCACCGACCTGCGGGCCGGACCTGGGCACTGCCGTGGC 1377

Qy   1015 actcgcctggctggagcaccgatgcgccgccttcgacgagtggagtggaactgacagcggc 1074
          |||||||||||| ||||||||| |||| |||||  |||||||| ||||||||||||||| ||
Db   1378 GCTCGCCTGGCTAGAGCACCGATGCGCGTCGCCTTTTGGCGAGTGGGAGTGGAACTGACAGCAGC 1437

Qy   1075 caaggctgattgctggctgcgggcccagcactgccttgacctgcccggcct 1134
          ||||||||||||||||||||  |||||||| ||||||||  |||| |||| |
Db   1438 CAAGGCTGATTGCTGGCTGCTGGGGCCAGCAGCTGCCTTGATCTGGCCGCCCCT 1497

Qy   1135 caaggcgcagccgaggggctcttcctgctactggcgccactgggaccaaaacctgcagct 1194
          |||||||  ||||| ||||||||||| |||||||| ||||||||||||||   |||||||
Db   1498 CAAGGCCCGCAGCCCGGGGCTCTTCCTCCTGCTACTGCGCCACTGGGACCACCTGCAGCT 1557

Qy   1195 acacctgctgtacagcccagcgaacgtgaaggctgccccctgctcttgggctgg 1254
          ||||||||||| |||||||||| ||| |||||||||||| ||||| ||| |||||
Db   1558 ACACCTGCTGTGTGCTACAGCCCAGCAAACCGTGAAGGCTGCCCCTGCTCGTGCGTTGGGTTGG 1617

Qy   1255 cgccccaccaacacactcaagtacctgcgcgccagggctggctcttggtgtgggaaa 1314
          ||||||||| ||||||| |   | ||||| |||| ||| |||||||||||||||||||
Db   1618 CGCCCCACCCACCAACACACT--AGACACTGCCGCCGCCAGGACTGCCCGTTGGTGGCTGGGAAG 1675

Qy   1315 gtgtaggcggtgccagcctgtccccaactgcttcttactcctccctgggagccccttg 1374
          |||||||||||| ||||| |||||| ||| |||| |||||| ||||| ||||||| |||
Db   1676 GTGTAGGCGGTGGTGCCGAGCTGTCCCACTGTTCCCTACTCCCTCCCTAGNGCCCCTCTTG 1735

Qy   1375 ccccccaaaaagtgcctgcctgcgtctctcctcctgctctcacagcca 1434
          ||||||||  || ||||||||| |  |||| ||||  |||||| |||||
Db   1736 CCCCCCAAGAAGAGTGCCTGCCTGTGCCCTGTGCTCTCTCCCCGCTCTCCACCCCCACCCCACACCCTCCC 1795
```

```
Qy   1435  ctccatcctctgagctccctgcaacacagtggaagggtagagagccacagtccccaatc 1494
           ||||||||||||||||||||  |||  |||||||||||  ||||||||||||||||||||
Db   1796  CTCCATCCTCTGAGCTCCCCGCAGCACAGTGGAAGGGGAGAGAGCCACAGTCCCCAAATC 1855

Qy   1495  ctatgcaataaagtgc 1510
           ||||||||||||||||
Db   1856  CTATGCAATAAAGTGC 1871
```

```
RESULT   2
BC001200
LOCUS       BC001200              3319 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3356192, mRNA, partial cds.
ACCESSION   BC001200
           BC001200
           BC001200.1  GI:12654716
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3319)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (11-DEC-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 6 Row: g Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..3319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3356192"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             <1..3008
                     /codon_start=3
                     /product="unknown (protein for IMAGE:3356192)"
                     /protein_id="AAH01200.1"
                     /db_xref="GI:12654717"
                     /translation="KAGPAGSSWTVKLGTWGPEAGPSLTELCPLLSEPHQPHLMLEGG
                     SLSSABYEARVRARRDFQRLQRRDSDGDRQVWFLQRRFHKDILLNPVLALSFCPDLSS
                     KPGHLGTATRELLFLLDSSVAHKDAIVLAVKSLPPGTLINLAVPGTLVQPLFPESRP
                     CSDDAVQLICESIETLQYPSGPPDVLAALDWAVGQPOHRAYPRQLFLLTAASPMAATT
                     HRTLELMRWHRGTARCFSFGLGPTCHQLLQGLSALSRQAYFLRPGORLQPMLVQALR
                     KALEPALSDISVDWFVPDTVEALLTPREIPALYPGDQLLGYCSLFRVDGFRSRPPGGQ
                     EPGWQSSGGSVFPSPEEAPSAASPGTEPTGTSEPLGTGTVSAELSSPWAARDSEQSTD
```

```
ALTDPVTDGPNPSDTAIWRRIFQSSYIREQVLTHCSASPEPGPGSTGSSESPGSOG
PGSPEGSAPLEPPSOOGCRSLAWGEPAGSRSCPLPAPTPAPVKVGALSTEVLGRQHRA
ALAGRSLSSPPGRANQVPGRPRKPSLGAILDGPSPEPGQLGQGLLDDSGNLLSPAPMD
WDMLMEPPFLFTAVPPSGELAPPAVPPQAPRCHVVIRGLCGEQPMCWEVGVGLETLWG
PGDSQOPPSPVRRAAWDQALHRLTAASVVRDNEQOLALRGGAETTADRGHARRCWLRA
LQTSKVSSAPSCFTCPVAVDATREVLPGALQVCSSEPAEPPGTPPASHSHLDAAPLP
TVVYSKGLQRGSPAGAWDSDQNGNSKRALGDPATPTEGPRRPRPPPCRLSMGRRHKL
CSPDPGQANNSEGSDHDYLPLVRLQEAPGSFRLDAPFCAAVRISQERLCRASPFAVHR
ASLSPTSASLPWALLGPGVGGQGDSATASCSPSPSSGSEGPGQVDSGRGSDTEASEGAE
GLGGTDLRGRTWATAVALAWLEHRCAAAFDEWELTAAKADCWLRAQHLPDGLDLAALK
AARGLFLLLRHWDQNLQLHLLCYSPANV"
```

```
BASE COUNT      586 a   1146 c    978 g    609 t
ORIGIN

Query Match             84.0%;  Score 1268.8;  DB 9;  Length 3319;
Best Local Similarity   99.8%;  Pred. No. 3.5e-219;
Matches 1270;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    239  ggccatgcccggaggtgctggcttcagcccttcagacacaagtaaggtcagctctgcccccc 298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2019  GGCCATGCCCGGAGGTGCTGGCTTCAGCCCTTCAGACACAGCAAGTAAGGTCAGCTCTGCCCCC 2078

Qy    299  tctgctcacttgccctgtagctggatgctactactagggaggtcctgctggggcc 358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2079  TCCTGCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGGAGGTCCTGCTGGGGCC 2138

Qy    359  ctgcaggtgcgagctcagagcgccgctgagcccccaggaacccctcctgcctctcacagc 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2139  CTGCAGGTGTGCAGCTCAGAGCCCGCTGAGCCCCCAGGAACCCCTCCTGCCTCTCACAGC 2198

Qy    419  catctagatgcagctcctctgcccactgttgtctactctaaaggactcagagaggctct 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2199  CATCTAGATGCAGCTCCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCT 2258

Qy    479  ccagcaggcgcctggactcggaccaaaatgcaactccaagcgtgctttggggaccct 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2259  CCAGCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGACCCT 2318

Qy    539  gccactcccacggaaggtcctgcgcccaccctcccgtcctccctgtcggctcagcatg 598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2319  GCCACTCCCACGGAAGGTCCTGCGCCCACCCTCCCGTCCTCCCTGTCCGGCTCAGCATG 2378

Qy    599  ggccgccgtcacaaactctgtagccctgaccccgggccaggccaacaaacagtgaaggcagc 658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2379  GGCCGCCGTCACAAACTCTGTAGCCCTGACCCCGGGCCAGGCCAACAAACAGTGAAGGCAGC 2438

Qy    659  gaccatgactacctgccctggtgcggctgcaggaggcaccaggcgctccttccgcctggac 718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2439  GACCATGACTACCTGCCCTGGTGCGGCTGCAGGAGGCACCAGGCGCTCCTTCCGCCTGGAC 2498

Qy    719  gcgcccttctgcgccggctgcgcatctgcgaggagcgcctctgcgctcgcgcctttt 778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2499  GCGCCCTTCTGCGCCGCGTGCGCCGTGTGCGCCATCTGCGAGGAGCGCCTCTGCGCTCGCGCCTT 2558

Qy    779  gccgtgcaccgcccagcctcagcccaccccagcctcggcccattgccctggcacttctgggc 838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2559  GCCGTGCACCGCGCCAGCCTCAGCCCACCTCGGCCTCCGGCCTCCATTGCCCTGGCACTTCTGGGC 2618

Qy    839  cctgtgttggccaggtgacagtgccacggcctcctgcagcccgtccccccagctcgggc 898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2619  CCTGGTGTTGGCCAGGGTGACAGTGCCACGGCCTCCTGCAGCCCGTCCCCCAGCTCGGGC 2678

Qy    899  tctgaggggccaggccaggtgacagtgggcggggctcagacaccgaggcctccgatggg 958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2679  TCTGAGGGGCCAGGCCAGGTGGACAGTGGGCGGGGCTCAGACACCGAGGCCTCCGAGGGG 2738

Qy    959  gcggaagggctgggcggccaccgacctgcggggcccggacctggcccactgccgtagcactc 1018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2739  GCGGAAGGGCTGGGCGGCGCCACCGACCTGCGGGGCCGGACCTGGGCCCACTGCCGTAGCACTC 2798

Qy   1019  gcctggctggagcaccaatgcgccgcgctccttcgacgagtgggaactgacagccggccaag 1078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2799  GCCTGGCTGGAGCACCGATGCGCCGCCTTCGCAGAGTGGGAACTGACAGCGGCCAAG  2858
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1079  gctgattgctgctgcgggcccagcacttgcctgacggccttgacctggccgccctcaag  1138
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2859  GCTGAGTTGCTGCGTCGCGGGCCCAGCACTTGCCTGACGGCCTTGACCTGGCCGCCCTCAAG  2918
Qy   1139  gccgcagcccagggggctcttcctgctactggcgccactggaccaaaacctgcagctacac  1198
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2919  GCCGCAGCCCGAGGGCGTCTTCCTGCTACTGCGCCACTGGGACCAAAACCTGCAGCTACAC  2978
Qy   1199  ctgctgctacagcccgaacgtgtgaaggctgcccctgctgttgggctgggcgcc  1258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2979  CTGCTGTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCTGCTGTTGGGCTGGGCGGCC  3038
Qy   1259  ccacccaacacactcaagtcactgcgcgcccaggggtcgcctcttggtgctgggaaagtgt  1318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3039  CCACCCAACACACTCAAGTCACTGCGCGCCCAGGGCTCGCCTCTTGGTGCTGTGGGAAAGTGT  3098
Qy   1319  aggctggtgcagcgctgtcccccactgcttcttactccctcggaggccctcttgcccc  1378
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3099  AGGCTGGTGCAGCCGTGCCAGCCCGTGCCCCCACTGCTCTCTTACTCCCTCCCTAGAGCCCTCTTGCCCC  3158
Qy   1379  cacaaaaagtgctgctgtctcctccctctctccaccccaccactcacactccccctcc  1438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3159  CACAAAAAGTGCTGCTGCTGTGCTCTCCTCCCTCTCTCCCTCCCACCCCACTCACACTCCCCCTCC  3218
Qy   1439  atcctctgagctccctgcaacacagtggagggtagagagccacagtcccaaatcctat  1498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3219  ATCCTCTGAGCTCCCTCCAACACAGTGGAAGGGTAGAGACGCCACAGTCCCCAAATCCTAT  3278
Qy   1499  gcaataaagtgc  1510
           ||||||||||||
Db   3279  GCAATAAAGTGC  3290
```

```
RESULT  3
AC061705
LOCUS       AC061705              151032 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-488M12, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC061705
VERSION     AC061705.14  GI:18449515
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151032)
AUTHORS     Muzny,D.M., Adams,C., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,H., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151032)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:9743372.
            -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: HANK
            Center clone name: RP11-488M12
            -------------- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy: 18% of reads
            Chemistry: Dye-terminator Big Dye: 82% of reads
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            Consensus quality: 140075 bases at least Q40
            Consensus quality: 148512 bases at least Q30
            Consensus quality: 153337 bases at least Q20
            Estimated insert size: 149429; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 16 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *            1: contig of 44630 bp in length
          *        44631: gap of unknown length
          *        44731: contig of 20614 bp in length
          *        65345: gap of unknown length
          *        65445: contig of 9916 bp in length
          *        75361: gap of unknown length
          *        75461: contig of 9166 bp in length
          *        84627: gap of unknown length
          *        84727: contig of 8697 bp in length
          *        93424: gap of unknown length
          *        93524: contig of 7308 bp in length
          *       100832: gap of unknown length
          *       100932: contig of 6830 bp in length
          *       107762: gap of unknown length
          *       107862: contig of 7418 bp in length
          *       115280: gap of unknown length
          *       115380: contig of 5216 bp in length
          *       120596: gap of unknown length
          *       120696: contig of ...
```

```
*    120696   128475: contig of 7780 bp in length
*    128476   128575: gap of unknown length
*    128576   134874: contig of 6299 bp in length
*    134875   134974: gap of unknown length
*    134975   140581: contig of 5607 bp in length
*    140582   140681: gap of unknown length
*    140682   143980: contig of 3299 bp in length
*    143981   144080: gap of unknown length
*    144081   146393: contig of 2313 bp in length
*    146394   146493: gap of unknown length
*    146494   148710: contig of 2217 bp in length
*    148711   148810: gap of unknown length
*    148811   151032: contig of 2222 bp in length.

FEATURES      Location/Qualifiers
     source   1..151032
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="3"
              /clone="RP11-488M12"

BASE COUNT   37567 a   38844 c   38131 g   35347 t   1543 others
ORIGIN
```

```
Query Match    80.6%;   Score 1217.8;  DB 2;   Length 151032;
Best Local Similarity  85.2%;  Pred. No. 3.9e-210;
Matches 1508;   Conservative   0;  Mismatches  2;   Indels 259;  Gaps  3;

Qy       1 tccagagtgctgaatttctggcagccaggggctcttgctctgctatgggttgaagatt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53444 TCCAGAGTGCTGAATTTCTGGCAGCCAGGGGCTCTTGCTCTGCTATGGGTTGAAGATT 53503

Qy      61 cagcctgatcacttcttgtcaaagttgggagtctgagctggttctgggcctagcaaggc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53504 CAGCCTGATCACTTCTTGTCAAAGTTGGGAGTCTGAGCTGGTTCTGGGCCTAGCAAGGC 53563

Qy     121 aatctactctgttaaagtcgatagaggagaagtggggagaacactgcaatttttctta 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53564 AATCTACTCTGTTAAAGTCGATAGAGGAGAAGTGGGGAGAACACTGCAATTTTTCTA 53623

Qy     181 tgagccctgtagagggtggtgtgtgggggctccacagactgctccttggttccacag 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53624 TGAGCCCTCTGTAGAGGGTGGTGGTGGGGGGCTCCACAGACTGCTCCTTGGTTCCACAG 53683

Qy     241 ccatgccggagtgcttcgagcctcagacaacaagtaaggtctgccctgccccctc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53684 CCATGCCGGAGTGCTGGCTTCGAGCCCCTTCAGACACAAGTAAGGTCAGCTCTGCCCCCTC 53743

Qy     301 ctgcttcactggccctgtagtctggatgctactactagggaggtcggctgctctgggacc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53744 CTGCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGAGGTCGGCTGCTCTGGGACC 53803

Qy     361 gcaggtgtgcagct 374
           ||||||||||||||
Db   53804 GCAGGTGTGCAGCTACACGTAGGGGCTCTCTTCTGTGACCTCTCAGGTGCAGCTCTCACCT 53863

Qy     375 -------------cagagcccgct 385
                        ||||||||||||
Db   53864 CACTGGGCCACTCTCCTGGACAAACCTGACTCCTTGTCCCATGTGCCCCTGCAGAGCCCGCT 53923

Qy     386 gagcccaggaaccccctccttcgcctctcacagccatctagatgcagctcctctgccact 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53924 GAGCCCCAGGAACCCCCTCCTTCGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACT 53983

Qy     446 gttgtcacttcaa 459
           |||||||||||||
Db   53984 GTTGTCACTTCTAAAGGTAACACCCAAAGGTAGGGAAAAGGGTAGGGCACTTGGGCTTAG 54043

Qy     460 -------------aggacttcagagaggctctccagcagg 486
                        ||||||||||||||||||||||||||
Db   54044 AGACCCGCCCCTGGCACTGATGATCCCCACTGACGACTTCAGGAGGCTCTCCAGCAGG 54103

Qy     487 cgcctgggactcggaccaaaatggcaactccaagccgtgcttgggggacctgccactcc 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55184 CAGTCCCCAAATCCTATGCAATAAAGTGC 55212
```

```
Db   54104 CGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCCGTGCTTGGGGGACCTGCCCACTCC 54163

Qy     547 cacggaaggtcctcgcgccactcccgtcctccccgtccccgtctcggctcagcatgggcgccg 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54164 CACGGAAGGTCCTCGCGCCACTCCCGTCCTCCCCGTCCCGTCTCGGCTCAGCATGGGCGCCG 54223

Qy     607 tcacaaactctgtagccctgaccgggccaacaacagtgaaggcagcgaccatga 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54224 TCACAAACTCTGTAGCCCTGACCCGGGCCAACAACAGTGAAGGCAGCGACCATGA 54283

Qy     667 ctacctgccctt------ 678
           ||||||||||||
Db   54284 CTACCTGCCCTTGGTGAGGACTCGGAGGTGGTGCCGCCGGGGCCCGGGGCGCTG 54343

Qy     679 ----------------ggtgcggctgcaggaggcaccag 701
                           |||||||||||||||||||||||
Db   54344 TTTCAGCTGCTTCTCCCCCCACCTCTCCTCTCCCAGGTGCCGGCTGCAGGAGGCACCAG 54403

Qy     702 gctccttccgcctggacgcgcccttcgccctctgcgccgctgtgcgactccgaggacgcctct 761
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54404 GCTCCTTCCGCCTGGACGCGCCCTTCGCCCTCTGCGCCGCTGTGCGCGACTCCGAGGACGCCCTCT 54463

Qy     762 gccgtgcctcgccctttgccgtcgcaccgcgcagcctcagcccacctcggcctcattgc 821
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54464 GCCGTGCCTCGCCCTTTGCCGTCGCACCGCGCAGCCTCAGCCCACCTCGGCCTCATTGC 54523

Qy     822 cctgggcacttctggccctggtgttggccaggtgacagtgcaacggcctcctgcgcagcc 881
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54524 CCTGGGCACTTCTGGCCCTGGTGTTGGCCAGGTGACAGTGCAACGGCCTCCTGCAGCC 54583

Qy     882 cgtcccccagctcggctcggggctctcgggggctctgcggggctctcagaca 941
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54584 CGTCCCCCAGCTCGGCTCTCGGGGCTCTCGGGGCTCTGCGGGGCTCTCAGACA 54643

Qy     942 ccgaggccctccgatgggcggaagggccggcagggacctgcgggggccggacctggg 1001
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54644 CCGAGGCCTCCGATGGGCGGAAGGGCTGGGCGGCACCGACCTGCGGGGACCTGGG 54703

Qy    1002 ccactgccgtagcactcgcctggctggagcaccgatgcgccgcctgccctgcgacggccttg 1061
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54704 CCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGCCGCCTGCCTTGCGACGGCCTTG 54763

Qy    1062 aactgacagcggccaaggctgattgctggctggcccgagggctcttcctgctactggcacctggaac 1121
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54764 AACTGACAGCGGCCAAGGCTGATTGCTGGCTGGCCCGAGGGCTCTTCCTGCTACTGGCCACTGGAAC 54823

Qy    1122 acctggccgcctcaaggccgcagcccgagggtctctccctgctcctgctcacctggac 1181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54824 ACCTGGCCGCCTCCAAGGCCGCAGCCCGAGGGTCTCTCCCTGCTCCTGCTCACTGGGACC 54883

Qy    1182 aaaacctgcagctacacctgctgtgtgctacagccccagcgaacgtgtgaaggctgccccctg 1241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54884 AAAACCTGCAGCTACACCTGCTGTGTGCTACAGCCCCAGCGAACGTGAAGGCTGCCGCCCTG 54943

Qy    1242 ctgcttgggctggcgcccaccccaccaaactcaagtcactgcgccccaggctggcctct 1301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54944 CTGCTTGGGCTGGCGGCCCCCCACCCCACCCAACTCAAGTCACTGCGCCCCAGGCTGGCGCTCT 55003

Qy    1302 tggtgctgggaaagtgtaggctggtgtgcagcctgtcctctccactgcttcttactcctccc 1361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55004 TGGTGCTGGGAAAGTGTAGGCTGTGCCAGCTGTGCCAGCTTCTTCTTACTCCTCCC 55063

Qy    1362 tggagccctctgccccacaaaaagtgctgctgcctgtctctcctcctcctccaccc 1421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55064 TAGAGCCCTCTGCCCCCACCACAAAAAGTGCCTGCCTGTGTCTCTCCTCCTCCTCCACCC 55123

Qy    1422 cactcacactcccctccatcctcgagctcctgagctccctgcaacacagtggaagagagca 1481
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55124 CACTCACACTCCCCTCCATCCTCTGAGCTCCCTGCAACACAGTGGAAGAGAGCCA 55183

Qy    1482 cagtcccaaatcctatgcaataagtgc 1510
           |||||||||||||||||||||||||||
Db   55184 CAGTCCCCAAATCCTATGCAATAAAGTGC 55212
```

RESULT   4
AC048331
LOCUS   AC048331   229586 bp   DNA   linear   HTG 31-JAN-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-48I02, WORKING DRAFT SEQUENCE,
            23 unordered pieces.
ACCESSION   AC048331
VERSION   AC048331.31 GI:18449516
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 229586)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,P., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,I., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,J., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,R., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,J., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 229586)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (14-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jan 31, 2002 this sequence version replaced gi:17974902.
            -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: HANQ
            Center clone name: RP11-48I02
            -------------- Summary Statistics

            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy: 30% of reads
            Chemistry: Dye-terminator Big Dye: 70% of reads
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 230423 bases at least Q40
                Consensus quality: 245348 bases at least Q30
                Consensus quality: 254396 bases at least Q20
                Estimated insert size: 236955; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length.
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 23 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            *   This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *   1   102971: contig of 102971 bp in length
            *   102972   103071: gap of unknown length
            *   103072   138452: contig of 35381 bp in length
            *   138453   138552: gap of unknown length
            *   138553   157218: contig of 18666 bp in length
            *   157219   157318: gap of unknown length
            *   157319   165594: contig of 8276 bp in length
            *   165595   165694: gap of unknown length
            *   165695   170673: contig of 4979 bp in length
            *   170674   170773: gap of unknown length
            *   170774   177654: contig of 6881 bp in length
            *   177655   177754: gap of unknown length
            *   177755   181770: contig of 4016 bp in length
            *   181771   181870: gap of unknown length
            *   181871   185296: contig of 3426 bp in length
            *   185297   185396: gap of unknown length
            *   185397   189950: contig of 4554 bp in length
            *   189951   190050: gap of unknown length
            *   190051   193028: contig of 2978 bp in length
            *   193029   193128: gap of unknown length
            *   193129   195613: contig of 2485 bp in length
            *   195614   195713: gap of unknown length
            *   195714   199136: contig of 3423 bp in length
            *   199137   199236: gap of unknown length
            *   199237   202491: contig of 3255 bp in length
            *   202492   202591: gap of unknown length
            *   202592   205157: contig of 2566 bp in length
            *   205158   205257: gap of unknown length
            *   205258   207869: contig of 2612 bp in length
            *   207870   207969: gap of unknown length
            *   207970   210484: contig of 2515 bp in length
            *   210485   210584: gap of unknown length
            *   210585   213846: contig of 3262 bp in length
            *   213847   213946: gap of unknown length
            *   213947   217298: contig of 3352 bp in length
            *   217299   217398: gap of unknown length
            *   217399   220482: contig of 3084 bp in length
            *   220483   220582: gap of unknown length
            *   220583   223048: contig of 2466 bp in length
            *   223049   223148: gap of unknown length
            *   223149   225284: contig of 2136 bp in length
            *   225285   225384: gap of unknown length
            *   225385   227396: contig of 2012 bp in length
            *   227397   227496: gap of unknown length
            *   227497   229586: contig of 2090 bp in length.
FEATURES   Location/Qualifiers
     source   1..229586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-48I02"

```
BASE COUNT    56229 a   57475 c   54442 g   59219 t   2221 others
ORIGIN

Query Match      80.6%;   Score 1217.8;  DB 2;  Length 229586;
Best Local Similarity   85.2%;  Pred. No. 3.7e-210;
Matches 1508;  Conservative  0;  Mismatches  2;  Indels 259;  Gaps 3;

Qy      1 tccagagtgctgaatttctgggcagccaggggtcttgctctgctatgggttgaagatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  83695 TCCAGAGTGCTGAATTTCTGGGCAGCCAGGGGTCTTGCTCTGCTATGGGTTGAAGATT 83754

Qy     61 cagcctgatacacttcttgtcaaagttgggagtctgagctggttctggcctagccaaggc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  83755 CAGCCTGATCACTTCTTGTCAAAGTTGGGAGTCTGAGCTGGTTCTGGCCTAGCCAAGGC 83814

Qy    121 aatctactctgttaaagtcgatagagggagagagctgggaagaacactgcaattttttcta 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  83815 AATCTACTCTGTTAAAGTCGATAGAGGGAGAGAGCTGGGAGGAACACTGCAATTTTTTCTA 83874

Qy    181 tgagccctgtagagggtgggtggtggtggggctccacagactgctccttggttccacagg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  83875 TGAGCCCTGTAGAGGGTGGGTGGTGGTGGGGGCTCCACAGACTGCTCCTTGGTTCCACAGG 83934

Qy    241 ccatgcccgaggtgctggcttcagacaagtaaggtcagctctgccccctc 300
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  83935 CCATGCCCGAGGTGCTGGCTTCGCGGGCCCAGCACTTGCCTGACGGCCTC 83994

Qy    301 ctgcttcacttgcctgtagctgtggatgctgtactactagggggagtcctgcctgggcct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  83995 CTGCTTCACTTGCCTGTGTAGCTGTGATGCTGTACTACTAGGAGGTCCTGCCTGCGGGCCT 84054

Qy    361 gcaggtgtgcagct 374
          ||||||||||||||
Db  84055 GCAGGTGTGCAGCTCAGGTAGGGGCTCTTCGTCTGGTGACCTCTCAGTGCAGCTCTCACCT 84114

Qy    375 ------------------------cagagcccgct 385
                                  ||||||||||||
Db  84115 CACTGGCCACTCTCTGGACAAACCTGACTCCGTCCCATCTGCCCCTGCAGAGCCCGCT 84174

Qy    386 gagcccccaggaagaccccctcctgcctctcacagccatctagatgcagctcctctgccact 445
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84175 GAGCCCCCAGGAACCCCTCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACT 84234

Qy    446 gttgtctactctaa 459
          ||||||||||||
Db  84235 GTTGTCTACTCTAAAGGTAACACCCAAAGGTAGGGAAAGGTAGGGCACTTGGGCTTAG 84294

Qy    460 ------------aggacttcagagaggctctccagcagg 486
                      ||||||||||||||||||||||||||
Db  84295 AGACCGCCCCCTGGCACTGATGATCCCCACTGCAGGACGCTCTCAGCAGG 84354

Qy    487 cgcctggactggaccaaaatggcaactccaagcgtgctttggggaccctgccactcc 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84355 CGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTCGTTTGGGGACCCTGCCACTCC 84414

Qy    547 cacggaaggtcctgcgccgccacctcccgtcctccctgtcggctcagcatggccgccg 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84415 CACGGAAGGTCCTCGCGCCGCCACCTCCCGTCCTCCCCGTCGGCCTCAGCATGGGCCGCCG 84474

Qy    607 tcacaaactctgtagccctgaccgggccaggccaacaacagtgaaggcagcgaccatga 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84475 TCACAAACTCTGTAGCCCTGACCCGGCCAGCCCAGCCCAACAACAGTGAAGGCAGCGACCATGA 84534

Qy    667 ctacctgccctt 678
          ||||||||||||
Db  84535 CTACCTGCCCTTGGTGAGGACTCGGGAGGTGGAGGTGGTGCCGCCGGGGCCGGGCGCTG 84594

Qy    679 ------------ggtgcggctgcaggaggaggcaccag 701
                    |||||||||||||||||||||||||
Db  84595 TTTCAGCTCGGTTCTCCCCACCTCCTCCTCTCCTCCAGGTGCGGCTGCAGGTGGCCACCAG 84654

Qy    702 gctccttccgccgacgcgcccttctgcgccgctgtgccgcatctcgcaggagcgcctct 761
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84655 GCTCCTTCGCCTGGACGCGCCCTTCTGCGCCGCTGCCGCATCCGCCGTGCCCACCTGGCCTCCATTGC 84714

Qy    762 gccgtgcctgccctttgccgtgaccgcgccagcctcagcccaccctcagcccacctggcctccattgc 821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84715 GCCGTGCCTGCCGCCCTTGCCGTGACCGCGCCAGCCTCAGCCCACCTCAGCCCACCTGGCCTCCATTGC 84774

Qy    822 cctggcacttctgggccctggtgttggcaggtgacagtgcacggcctcctgcagcc 881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84775 CCTGGGCACTTCTGGGCCCTGGTGTGGGCAGGTGACAGTGCACGGCCTCCTGCAGCC 84834

Qy    882 cgtccccagctcgggctctgaggggccagtgcagtggacagtggcggggctcagaca 941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84835 CGTCCCCAGCTCGGGTCTCGAGGGGCCAGTGCAGTGGACAGTGGACAGTGGGCGGGGCTCAGACA 84894

Qy    942 ccgaggcctcgatgtgggcggaagggctgggcggcaccgacctgcggggccggacctggg 1001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84895 CCGAGGCCTCCGAGGGGCCGGAGGGGCCGGAAGGGCTGGGCGGCACCGACCTGCGGGGCCGGACCTGGG 84954

Qy   1002 ccactccgtagcactcgcctggctggagcaccgatggcgccgctgccttgccctgacgagtggg 1061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84955 CCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGGCGCCGGCCGCCTGCCTTCGACGAGTGGG 85014

Qy   1062 aactgacagcggccaaggctgattgctggctgcgggcccagcacttgcctgctgacggccttg 1121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85015 AACTGACAGCGGCCAAGGCTGATTGCTGGCTGCGGGCCCAGCACTTGCCTGCTGACGGCCTTG 85074

Qy   1122 acctgccgccctcaaggccgcagcccgaggggctcttcctgctactggcgccactggacc 1181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85075 ACCTGGCGCCCTCAAGGCCGCAGCCCGCCTCAAGGCCTCTCCTCCTACTACTGCGGCCACTGGGACC 85134

Qy   1182 aaaacctgcagctacacctgtgtgctacagccagcgaacgtgtgaaggctgcccctg 1241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85135 AAAACCTGCAGCTACACCTGCGTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCTG 85194

Qy   1242 ctgcttgggctggcgcccaccaacactcaagtcactccgcccaggggctggcctct 1301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85195 CTGCTTGGGCTGGGCCGCCCACCACCCACCAACACTCAAGTCACTGCCCCCAGGGGCTGGCCTCT 85254

Qy   1302 tggtgctggaaagtgtaggctggtgccagcctgtccccccactgcttctttactccctccc 1361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85255 TGGTGCTGGGGAAAGTGTAGGCTGGTGGTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCC 85314

Qy   1362 tggagccctcttgtgcccacaaaaagtgctgctgcgtcgtgtcctctccctcctccaccc 1421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85315 TAGAGCCCTCTTGTGCCCCCACCAAAAAGTGCCTCTGCCGTGCCTGTCTGCTCTCCCTCCTCCCACCCC 85374

Qy   1422 cactcacactcccctccctcccatcctctgagctcctgcaacacagtggaagggtagagagcca 1481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85375 CACTCACACTCCCCTCCCTCCATCCTCTGAGCTCCTGCAACACAGTGGAAGGGTAGAGAGCCA 85434

Qy   1482 cagtcccaaatcctatgcaataaagtgc 1510
          ||||||||||||||||||||||||||||
Db  85435 CAGTCCCAAATCCTATGCAATAAAGTGC 85463
```

```
RESULT 5
AC024530/c
LOCUS       AC024530              152623 bp    DNA     linear   HTG 14-MAR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-131020 map 4, *** SEQUENCING
            IN PROGRESS ***, 49 unordered pieces.
ACCESSION   AC024530
VERSION     AC024530.3  GI:7239669
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152623)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 4, clone RP11-131020
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 152623)
```

AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
           Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
           DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
           Fenestor,C., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
           Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
           Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,M., McCarthy,M.,
           McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
           Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
           Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
           Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
           Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
           Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
           Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
           Zody,M.
LE
RNAL   Direct Submission
           Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 14, 2000 this sequence version replaced gi:7229935.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ---------- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           ---------- Project Information
           Center project name: L7361
           Center clone name: 131_O_20
           ----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 49 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *   1   1156: contig of 1156 bp in length
           *   1157   1256: gap of   100 bp
           *   1257   2440: contig of 1184 bp in length
           *   2441   2540: gap of   100 bp
           *   2541   3807: contig of 1267 bp in length
           *   3808   3907: gap of   100 bp
           *   3908   5016: contig of 1109 bp in length
           *   5017   5116: gap of   100 bp
           *   5117   6229: contig of 1113 bp in length
           *   6230   6329: gap of   100 bp
           *   6330   7466: contig of 1137 bp in length
           *   7467   7566: gap of   100 bp
           *   7567   8835: contig of 1269 bp in length
           *   8836   8935: gap of   100 bp
           *   8936   10247: contig of 1312 bp in length
           *   10248   10347: gap of   100 bp
           *   10348   11428: contig of 1081 bp in length
           *   11429   11528: gap of   100 bp
           *   11529   12880: contig of 1352 bp in length
           *   12881   12980: gap of   100 bp
           *   12981   13991: contig of 1011 bp in length
           *   13992   14091: gap of   100 bp
           *   14092   15759: contig of 1668 bp in length
           *   15760   15859: gap of   100 bp
           *   15860   17156: contig of 1297 bp in length
           *   17157   17256: gap of   100 bp
           *   17257   18728: contig of 1472 bp in length
           *   18729   18828: gap of   100 bp
           *   18829   20169: contig of 1341 bp in length
           *   20170   20269: gap of   100 bp
           *   20270   21411: contig of 1142 bp in length
           *   21412   21511: gap of   100 bp
           *   21512   22964: contig of 1453 bp in length
           *   22965   23064: gap of   100 bp
           *   23065   25024: contig of 1960 bp in length
           *   25025   25124: gap of   100 bp
           *   25125   27388: contig of 2264 bp in length
           *   27389   27488: gap of   100 bp
           *   27489   28881: contig of 1393 bp in length
           *   28882   28981: gap of   100 bp
           *   28982   30874: contig of 1893 bp in length
           *   30875   30974: gap of   100 bp
           *   30975   33014: contig of 2040 bp in length
           *   33015   33114: gap of   100 bp
           *   33115   34935: contig of 1821 bp in length
           *   34936   35035: gap of   100 bp
           *   35036   37222: contig of 2187 bp in length
           *   37223   37322: gap of   100 bp
           *   37323   39483: contig of 2161 bp in length
           *   39484   39583: gap of   100 bp
           *   39584   41860: contig of 2277 bp in length
           *   41861   41960: gap of   100 bp
           *   41961   44448: contig of 2488 bp in length
           *   44449   44548: gap of   100 bp
           *   44549   46452: contig of 1904 bp in length
           *   46453   46552: gap of   100 bp
           *   46553   49889: contig of 3337 bp in length
           *   49890   49989: gap of   100 bp
           *   49990   52994: contig of 3005 bp in length
           *   52995   53094: gap of   100 bp
           *   53095   56588: contig of 3494 bp in length
           *   56589   56688: gap of   100 bp
           *   56689   59956: contig of 3268 bp in length
           *   59957   60056: gap of   100 bp
           *   60057   63904: contig of 3848 bp in length
           *   63905   64004: gap of   100 bp
           *   64005   67438: contig of 3434 bp in length
           *   67439   67538: gap of   100 bp
           *   67539   70684: contig of 3146 bp in length
           *   70685   70784: gap of   100 bp
           *   70785   75479: contig of 4695 bp in length
           *   75480   75579: gap of   100 bp
           *   75580   79243: contig of 3664 bp in length
           *   79244   79343: gap of   100 bp
           *   79344   82983: contig of 3640 bp in length
           *   82984   83083: gap of   100 bp
           *   83084   87910: contig of 4827 bp in length
           *   87911   88010: gap of   100 bp
           *   88011   94315: contig of 6305 bp in length
           *   94316   94415: gap of   100 bp
           *   94416   99313: contig of 4898 bp in length
           *   99314   99413: gap of   100 bp
           *   99414 104977: contig of 5564 bp in length
           * 104978 105077: gap of   100 bp
           * 105078 112571: contig of 7494 bp in length
           * 112572 112671: gap of   100 bp
           * 112672 118896: contig of 6225 bp in length
           * 118897 118996: gap of   100 bp
           * 118997 125479: contig of 6483 bp in length
           * 125480 125579: gap of   100 bp
           * 125580 130800: contig of 5221 bp in length
           * 130801 130900: gap of   100 bp
           * 130901 137670: contig of 6770 bp in length
           * 137671 137770: gap of   100 bp
           * 137771 145779: contig of 8009 bp in length
           * 145780 145879: gap of   100 bp
           * 145880 152623: contig of 6744 bp in length.
FEATURES   Location/Qualifiers
     source   1..152623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
              /chromosome="4"
              /map="4"
              /clone="RP11-131020"
              /clone_lib="RPCI-11 Human Male BAC"
misc_feature  1..1156
              /note="assembly_fragment"
misc_feature  1257..2440
              /note="assembly_fragment"
misc_feature  2541..3807
              /note="assembly_fragment"
misc_feature  3908..5016
              /note="assembly_fragment"
misc_feature  5117..6229
              /note="assembly_fragment"
misc_feature  6330..7466
              /note="assembly_fragment"
misc_feature  7567..8835
              /note="assembly_fragment"
misc_feature  8936..10247
              /note="assembly_fragment"
misc_feature  10348..11428
              /note="assembly_fragment"
misc_feature  11529..12880
              /note="assembly_fragment"
misc_feature  12981..13991
              /note="assembly_fragment"
misc_feature  14092..15759
              /note="assembly_fragment"
misc_feature  15860..17156
              /note="assembly_fragment"
misc_feature  17257..18728
              /note="assembly_fragment"
misc_feature  18829..20169
              /note="assembly_fragment"
misc_feature  20270..21411
              /note="assembly_fragment"
misc_feature  21512..22964
              /note="assembly_fragment"
misc_feature  23065..25024
              /note="assembly_fragment"

Query Match    80.4%;  Score 1214.6;  DB 2;  Length 152623;
Best Local Similarity 85.1%;  Pred. No. 1.4e-209;
Matches 1506;  Conservative   0;  Mismatches   4;  Indels  259;  Gaps    3;

Qy      1 tccagagtgctgaatttctgggcagcaggggctcttgctctgctatgggtgaagatt 60
          |||||||||||||||||||||||||| ||||||||||| |||| |||||||||||||
Db 142405 TCCAGAGTGCTGAATTTCTGGGCAGCCAGGGGCTCTTGCTCTGCTATGGGTTGAAGATT 142346

Qy     61 cagccgatcacttcttgtcaaagttgggaatctgagctggttctggcctagcaaggc 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142345 CAGCCCTGATCACTTCTTGTGTCAAAGTTGGGGAGTCTGGGTTCTGGGCCTAGCAAGGC 142286

Qy    121 aatctactctgttaaagtcgatagagggagaagctgggggagaacactgcaatttttcta 180
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 142285 AATCTACTCTGTTAAAGTCGATAGAGGGAGAGGAAGCTGGGGAGAACACTGCAATTTTTCTA 142226

Qy    181 tgagccctcgtagagggtgctggcttcgagcctccagactgctccttggttccacagg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142225 TGAGCCCTCGTAGAAGGGTGGGGTGGTGGGGGTCCCACAGACTGCTCCTTGGTTCCACAGG 142166

Qy    241 ccatgcccggagggtgctgcctgtcggatgctgctactaggagggaggtcctgctgggccct 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142165 CCATGCCCGGAGGGTGCTGCCTGCTCGCGTCCAGCCTTCAAACAGTAAGGTCAGCTCTGCCCCCTC 142106

Qy    301 ctgcttcacttgccctgtagctgtgggatgctgctgactacagggaggtcctgctgggccct 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142105 CTGCTTCACTTGCCCTGTAGCTGTGGATGCTGTGGATGCTACTACTAGGGAGGTCCTGCCTGCTGGGCCCT 142046

Qy    361 gcaggtgtgcagct----------------------------------------- 374
          ||||||||||||||
Db 142045 GCAGGTGTGCAGCTCAGGTAGGGCCTCTTCTCGTGACCTCTCAGGTGCAGCTCTCACCT 141986

Qy        -----------cagagcccgct 385
                     ||||||||||||
```

```
Db 141985 CACTGGCCACTCTCCTCTGGACAAACCTGACTCCTGTCCCATGTGCCCTGCAGAGCCCGCT 141926
                 |||||||||
Qy    386 gagcccccaggaaccccctcctgctctcacagccatctagatgcagctcctctgccact 445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141925 GAGCCCCCAGGAACCCCCTCCTGCTCTCACAGCCATTAGATGCAGCTCCTCTGCCCACT 141866

Qy    446 gttgctactctaa--------------------------------------------- 459
          |||||||||||||
Db 141865 GTTGTCTACTCTAAAGGTAACACCCAAAGGTAGGGAAAAGGTAGGGGCACTTGGGCTTAG 141806

Qy    460 ------------------------------------aggacttcagagagggctccagcagg 486
                                              ||||||||||||||||||||||||
Db 141805 AGACCCGCCCCTGCACTGCACTGATGATCCCCACTGCAGGACTTCAGAGAGGGCTCCAGCAGG 141746

Qy    487 cgcctgggactcggaccaaaaatggcaactccaagcgtgctttgggggacctgccactcc 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141745 CGCCTGGGACTCGGACCAAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCTGCCCACTCC 141686

Qy    547 cacggaaggtcctcgcgccacctcccgtcctccctgtggctcagcatggccgcg 606
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141685 CACGGAAGGTCCTCGCGCCACCTCCCGTCCTCCCTGTCGGCTCAGCATGGGCCGCCG 141626

Qy    607 tcacaaactctgtagccctgacccgggccaggccaacaacagtgaaggcagcgaccatga 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141625 TCACAAACTCTGTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGA 141566

Qy    667 ctacctgcccct------------------------------------- 678
          ||||||||||||
Db 141565 CTACCTGCCCTTGGTGTGAGGACTCGGAGGTGGTGCCGCGGGGCGCGCTG 141506

Qy    679 ---------------ggtgcggctgcaggaggcaccag 701
                         |||||||||||||||||||||||
Db 141505 TTTCAGCTCGCTTCTCCCCCACCTCCTCTCCCAGGTGCCGGCTGCAGGAGGCACCAG 141446

Qy    702 gctccttccgcccctgacgcgccctctcgcgctgtgcgcatctgcaggagcgcctct 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141445 GCTCCTTCCGCCCTGACGCGCCCTTCTGCGCCGTGCGCCGTGCTGCCATCTCGCAGGAGCGCCTCT 141386

Qy    762 gccgtgcctgcgcccttttgccgtgcaccgcgcagcctcagcctccagcccccacctcggcctcattgc 821
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141385 GCCGTGCCTGCGCCCTTTGCCGTGCACCGCGCCAGCCTCAGCCCCCACACCTCGGCCTCATTGC 141326

Qy    822 cctgggcacttctgggccctggtgttggccaggtgacagtgcacagcgcctcctgcagcc 881
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141325 CCTGGGCACTTCTGGGCCCTGGTGTTCGGCCAGGTGACAGTGCACGGCCCTCCTGCAGCC 141266

Qy    882 cgtccccagctcggggtcttgagggggccaggccaggccaggtggacagtggcgggtccagaca 941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141265 CGTCCCCAGCTCGGGGTCTCGGAGTCTGAGGGGGCCAGGGGGCCAGGTGGACAGTGGCGGGGCTCAGACA 141206

Qy    942 ccgaggcctccgatggggcgggaaggggctgggcggccaggccgacctggg 1001
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141205 CCGAGGCCTCCGAGGGGCGGGAAGGGCTGGGGCGGGGCCAGGCCGACCTGGG 141146

Qy   1002 ccactgccgtagcactcgcctggctgagcaccgatgcgccgctgccttcgacgagtggg 1061
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141145 CCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGGCGCCGCTGCCTTCGACGAGTGGG 141086

Qy   1062 aactgacagcggcaaggctgattgctggctgctcgcggcccagaggtcttcctgcctgacggccttg 1121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141085 AACTGACAGCGGCAAGGCTGATTGCTGGGCTGCTCGCGGCCCAGAGGTCTTCCTGCTACTGGGACC 141026

Qy   1122 acctggcccgccctcaaggccagcccagcccagcccagcccattgggaccactgggacc 1181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141025 ACCTGGCCCGCCCTCAAGGCCCAGCCCAGCCCAGCCCAGGGCTGCTACTGCGCCACTGGGAC 140966

Qy   1182 aaaacctgcagctacacctgctgtctacagcccagcgaacgtgtgaaggctgcccccctg 1241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140965 AAAACCTGCAGCTACACCTGCTGTCTGCTACAGCCCAGCTGCCGAACGTGTGAAGGCTGCCCCCCTG 140906

Qy   1242 ctgcttgggctgggcgcccaccccaacacactcaagtcactgcgccgccaggggctggcctct 1301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 140905 CTGCTTGGCTGGCGCCCCACCCACCACACTCAAGTCACTGCCGCCCAGGGCTGGCCTCT 140846
Qy   1302 tggtgctgggaaagtgtaggctggtggcagcctgtccccactgcttcttactccctccc 1361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140845 TGGTGCTGGGAAAGTGTAGGCTGGTGGCCAGCCTGTCCCCACTGCTTCTTACTCCCTCCC 140786
Qy   1362 tggagccctctgccccacaaaagtgcctgctgtctctcccctcctcccaccc 1421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140785 TGGAGCCCTCTTGCCCCCACAAAAAGTGCCTGCTGTCTCTCCCCTCCTCCCACCC 140726
Qy   1422 cactcaactcccctccatcctctgagctccctgcaacacagtggaagggtagagagcca 1481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140725 CACTCACACTCCCCTCCCATACTCTGAGCTCCCTGCAACACAGTGGAAGGGTAGAGAGCCA 140666
Qy   1482 cagtccccaaatcctatgcaataaagtgc 1510
          |||||||||||||||||||||||||||||
Db 140665 CAGTCCCCAAATCCTATGCAATAAAGTGC 140637
```

RESULT 6
092931/c

```
            AC092931      298408 bp    DNA     linear   HTG 09-AUG-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-131020, WORKING DRAFT
            SEQUENCE, 42 unordered pieces.
ACCESSION   AC092931
VERSION     AC092931.1  GI:15135474
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE   1 (bases 1 to 298408)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mel.G., Metzker,M., Miner,Z., Miner,V., Mitchell,T., Mohabbat,M.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,M., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,L., Usmani,K., Vasquez,J., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 298408)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   ------------ Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ------------ Project Information
            Center project name: HCXV
            Center clone name: RP11-131020
            ------------ Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 276956 bases at least Q40
            Consensus quality: 302770 bases at least Q30
            Consensus quality: 320294 bases at least Q20
            Estimated insert size: 310663; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 3.7x in Q20 bases; sum-of-contigs estimation
            ----
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
          * NOTE: This sequence may represent more than one clone.
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 42 contigs. The true order of the pieces
          *   is not known and their order in this sequence record is
          *   arbitrary. Gaps between the contigs are represented as
          *   runs of N, but the exact sizes of the gaps are unknown.
          *   This record will be updated with the finished sequence
          *   as soon as it is available and the accession number will
          *   be preserved.
          *   1   33307: contig of 33307 bp in length
          *   33308   33407: gap of unknown length
          *   33408   55576: contig of 22169 bp in length
          *   55577   55676: gap of unknown length
          *   55677   71524: contig of 15848 bp in length
          *   71525   71624: gap of unknown length
          *   71625   84811: contig of 13187 bp in length
          *   84812   84911: gap of unknown length
          *   84912   96841: contig of 11930 bp in length
          *   96842   96941: gap of unknown length
          *   96942   107204: contig of 10263 bp in length
          *   107205   107304: gap of unknown length
          *   107305   119046: contig of 11742 bp in length
          *   119047   119146: gap of unknown length
          *   119147   128812: contig of 9666 bp in length
          *   128813   128912: gap of unknown length
          *   128913   139236: contig of 10324 bp in length
          *   139237   139336: gap of unknown length
          *   139337   148308: contig of 8972 bp in length
          *   148309   148408: gap of unknown length
          *   148409   156373: contig of 7965 bp in length
          *   156374   156473: gap of unknown length
          *   156474   165327: contig of 8854 bp in length
          *   165328   165427: gap of unknown length
          *   165428   173514: contig of 8087 bp in length
          *   173515   173614: gap of unknown length
          *   173615   180753: contig of 7139 bp in length
          *   180754   180853: gap of unknown length
          *   180854   187919: contig of 7066 bp in length
          *   187920   188019: gap of unknown length
          *   188020   195551: contig of 7532 bp in length
          *   195552   195651: gap of unknown length
          *   195652   202233: contig of 6583 bp in length
          *   202234   202334: gap of unknown length
          *   202335   208703: contig of 6369 bp in length
          *   208704   208803: gap of unknown length
          *   208804   216638: contig of 7835 bp in length
          *   216639   216738: gap of unknown length
          *   216739   221125: contig of 4387 bp in length

```
*    221126  221225: gap of unknown length
*    221226  225147: contig of 3922 bp in length
*    225148  225247: gap of unknown length
*    225248  230808: contig of 5561 bp in length
*    230809  230908: gap of unknown length
*    230909  236270: contig of 5362 bp in length
*    236271  236370: gap of unknown length
*    236371  240940: contig of 4570 bp in length
*    240941  241040: gap of unknown length
*    241041  244908: contig of 3868 bp in length
*    244909  245008: gap of unknown length
*    245009  249141: contig of 4133 bp in length
*    249142  249241: gap of unknown length
*    249242  254190: contig of 4949 bp in length
*    254191  254290: gap of unknown length
*    254291  257741: contig of 3451 bp in length
*    257742  257841: gap of unknown length
*    257842  260681: contig of 2840 bp in length
*    260682  260781: gap of unknown length
*    260782  263548: contig of 2767 bp in length
*    263549  263648: gap of unknown length
*    263649  266637: contig of 2989 bp in length
*    266638  266737: gap of unknown length
*    266738  269758: contig of 3021 bp in length
*    269759  269858: gap of unknown length
*    269859  273130: contig of 3272 bp in length
*    273131  273230: gap of unknown length
*    273231  275969: contig of 2739 bp in length
*    275970  276069: gap of unknown length
*    276070  280320: contig of 4251 bp in length
*    280321  280420: gap of unknown length
*    280421  282684: contig of 2264 bp in length
*    282685  282784: gap of unknown length
*    282785  285710: contig of 2926 bp in length
*    285711  285810: gap of unknown length
*    285811  288736: contig of 2926 bp in length
*    288737  288836: gap of unknown length
*    288837  290085: contig of 2049 bp in length
*    290086  290185: gap of unknown length
*    290186  294160: contig of 3175 bp in length
*    294161  294260: gap of unknown length
*    294261  296277: contig of 2017 bp in length
*    296278  296377: gap of unknown length
*    296378  298408: contig of 2031 bp in length.

FEATURES          Location/Qualifiers
     source       1..298408
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="3q"
                  /clone="RP11-131O20"
BASE COUNT    77516 a  69113 c  67054 g  80533 t  4192 others
ORIGIN
```

```
Query Match          80.2%;  Score 1211.4;  DB 2;  Length 298408;
Best Local Similarity 85.0%;  Pred. No. 5.1e-209;
Matches 1504;  Conservative 0;  Mismatches 6;  Indels 259;  Gaps 3;

Qy         1 tccagagtgctgaatttctgggcagcagggggctccttgtcctgctatgggttgaagatt 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184561 TCCAGAGTGCTGAATTTCTGGGCAGCAGGGGGCTCCTTGTCCTGCTATGGGTTGAAGATT 184502

Qy        61 cagcctgatcacttcttgtcaaagttgggagtctgagctggttctgggcctagacaaggc 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184501 CAGCCTGATCACTTCTTGTCAAAGTTGGGAGTCTGAGCTGGTTCTGGGCCTAGACAAGGC 184442

Qy       121 aatctactctgttaaagtcgatagagggagaagctgggagaacactgcaattttttcta 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184441 AATCTACTCTGTTAAAGTCGATAGAGGGAGAAGCTGGGAGAACACTGCAATTTTTTCTA 184382

Qy       181 tgagccctctgtagagggtggtggtggggggtgtgggggctccacagactgtcctcttgt 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184381 TGAGCCCTCTGTAAAGGGTGGTGGTGGGGGCTGTGGGGGCCTCCCACAGACTGCTCCTTGGTTCCACAGG 184322
```

```
Qy       241 ccatgccggagtgctgctgcttcgagccttcgagcagccctcagacaagtaaggtcagtcctgccccctc 300
             |||||||||||     |||||||||||       |||||||    ||||     |||||||||
Db    184321 CCATGCCGGAGGTGCTTCGAGCCTTCAAACAGCAGTAAGCTAGGTCAGCTGCCCCCTC 184262

Qy       301 ctgcttcacttgcctgtagctgtgatgtactactcaggaggtcctgctgggggcct 360
             |||||||||||||        |||||||||||||||||||||||||||||||||||
Db    184261 CTGCTTCACTTGCCCTGTAGCTGTGGATGTGGATGTCACTACTAGGGAGGTCCTGCTGGGGGCCCT 184202

Qy       361 gcagtgtgcagct 374
             ||||||||||||
Db    184201 GCAGGGTGCAGCTCAGGTAGGGCCTCCTTCGTGACCTCTCAGTGACCTCACCT 184142

Qy       375 cagagcccgct 385
                          ||||||||||
Db    184141 CACTGGCCACTCTCTGGACAAACCTGACTCCGTCCCATGTGCCCCTGCAGAGCCCGCT 184082

Qy       386 gagccccaggaacccctgcctctcacagccatctagatgcagctcctctgcccact 445
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184081 GAGCCCCAGGAACCCCTGCCTCTCTGCCTCTCACAGCAGCCATATAGATGCAGCTCCTCTGCCCACT 184022

Qy       446 gttgtctactctaa 459
             |||||||||||
Db    184021 GTTGTCTACTCTAAAGGTAACACCCAAAGGTAGGGAAAAGGTAGGGCACTTGGGCTTAG 183962

Qy       460 aggactcagagagggctctccagcagg 486
                              |||||||||||||||
Db    183961 AGACCCCCCCTGGCACTGATGATCCCCACTGCAGGACTTCAGAGAGGGCTCTCCAGCAGG 183902

Qy       487 cgcctgggactcggaccaaaatgcaactcaagcgtgcttttgggggaccctgcactcc 546
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183901 CGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCTGCCACTCC 183842

Qy       547 cacggaagtcctcgccgccactcccgtcctccctgtcggctcagcatgggcgccg 606
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183841 CACGGAGGTCCTCGCCGCCACTCCCCGTCCTCCCTGTCGGCTCAGCATGGGCGGCCG 183782

Qy       607 tcacaaactctgtagccctgacccggccaggccgccaacaacagtgaaggcagcgaccatga 666
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183781 TCACAAACTCTGTAGCCCTGACCCGGCCAGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGA 183722

Qy       667 ctacctgccctt 678
             ||||||||||
Db    183721 CTACCTGCCCTTGGTGAGGACTCGGGAGTGGGAGGTGCCGCGGGGCCGGGGCGCTG 183662

Qy       679 ggtgcggctgcaggaggagcaccagg 701
                             ||||||||||||||||||||
Db    183661 TTTCAGCTGCGTTCTCCCCCACCTCCTCTCCTCCAGGTGCGGCTGCAGGAGGAGCACCAG 183602

Qy       702 gctccttccgctgacgccgccttgcgcgcgtgcgcatctgcagggagcgcctct 761
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183601 GCTCCTTCCGCTGACGCCGCCCTTCTCCGCGCCCGTGCGCGCCATCTGCAGGGAGCGCCTCT 183542

Qy       762 gccggcctgccctttgccgtgcacgcgcagcctcagcccaacctcggcctcattgc 821
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183541 GCCGTGCCTCCGCTCCCCTTCGCCGTGCACCGCCGCCAGCCTCAGCCCAACCTCGGCCTCATTGC 183482

Qy       822 cctgggcacttctgggccctggtgttggccaggtgacagtgcacggcctcctgcagcc 881
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183481 CCTGGGCACTTCTGGGCCCTGGTGTTGGCCAGGTGACAGTGCCAGGTGACAGTGCCACGGCCTCCTGCAGCC 183422

Qy       882 cgtccccagctcgggctcggggagctgagggtctgaggggctcagtagtgggcgggcgggctcagaca 941
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183421 CGTCCCCAGCTCGGGCTCGGGGAGCTGAGGGGCTCGAGGGGCAGGGTCGGGGCGGGCTCAGACA 183362

Qy       942 ccgaggcctccgatggggcggaagggctcggcggcaccgaccctcgggggccggacctggg 1001
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183361 CCGAGGCCTCCGAGGGGCGGAAGGGCTGGGCGGCCACCGACCCTCGGGGGCCGGACCTGGG 183302

Qy      1002 ccactgccgtagcactcgcctggctggaggcaccgatcgccgctgccttcagcagtggg 1061
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183301 CCACTGCCGTAGCACTCGCCTGGCTGGAGCACCCGATCGCCGCTGCCTTCGACGAGTGGG 183242
```

```
Qy   1062 aactgacagcggccaaggctgattgctggctgcgggccagcacttgctgacggcctg 1121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183241 AACTGACAGCGGCCAAGGCTGATTGCTGGCTGCGGGCCAGCACTTGCTGACGGCCTTG 183182

Qy   1122 actggccgccctcaaggccgcagcccgagggcttcctgctactggcgccactgggacc 1181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183181 ACCTGGGCGCCCCTCAAGGCCGCCGCAGCCCGAGGGCTTCTTCGCTACTGCGCCACTGGGACC 183122

Qy   1182 aaaacctgcagctacacctgctgtgctacagcccagcgaacgtgtgaaggctgcccctg 1241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183121 AAAACCTGCAGCTACACCTGCTGTGCTACACAGCCCCAGCGAACGTGTGAAGGCTGCCCCCTG 183062

Qy   1242 ctgcttgggctggcggccccaccaccaagtcactgccccagggctggcctct 1301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183061 CTGCTTGGGCTGGCGGCCCCACCACCAAGTCACTGCCCCCAGGGCTGGCCTCT 183002

Qy   1302 tggtgctggaaagtgtaggctggttccagcctgtccccactgcttcttactcctccc 1361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183001 TGGTGCTGGAAAGTGTAGGCTGGTTGGTCCAGCCTGCCCCCACTGCTTCTTACTTCCTCCC 182942

Qy   1362 tgagccctctggcccccacaaaaagtgcctggcctgtgtctctccctctcctccaccc 1421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182941 TGGAGCCCCTTCTGGCCCCCACAAAAAGTGCCTGGCCTGTGTCTGTCTCCCTCTCCTCCCACCC 182882

Qy   1422 cactcaactccctccctcatcctctgagctccctgcaacacagtgaagggtagagagca 1481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182881 CACTCAACTCCCTCCCTCCATACTCTGAGCTCCCTGCAACACAGTGGAAGGGTAGAGAGCCA 182822

Qy   1482 cagtccccaaatcctatgcaataaagtgc 1510
          |||||||||||||||||||||||||||||
Db 182821 CAGTCCCCAAATCCTATGCAATAAAGTGC 182793
```

```
RESULT   7
AB047829
LOCUS        AB047829              2814 bp    mRNA    linear    PRI 11-OCT-2001
DEFINITION   Macaca fascicularis brain cDNA, clone:QccE-10361.
ACCESSION    AB047829
VERSION      AB047829.1  GI:9967104
KEYWORDS     fis (full insert sequence).
SOURCE       Macaca fascicularis adult male cerebellum cortex cDNA to mRNA,
             clone_lib:macaque brain cDNA library QccE clone:QccE-10361.
  ORGANISM   Macaca fascicularis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
             Cercopithecinae; Macaca.
REFERENCE    1  (bases 1 to 2814)
  AUTHORS    Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirata,M.,
             Suto,Y., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
  TITLE      Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
             chromosomes
  JOURNAL    Gene 275 (1), 31-37 (2001)
```
```
REFERENCE    2  (bases 1 to 2814)
  AUTHORS    Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-AUG-2000) Katsuyuki Hashimoto, National Institute of
             Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
             1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
             (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,
             Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
COMMENT      Lab host:       TOP10
             Vector:         pME18S-FL3 (Acc.No. AB009864)
             R. Site1:       DraIII (CACTGTGTG)
             R. Site2:       DraIII (CACCATGTG)
             Description: 1st strand cDNA was primed with an oligo(dT) primer
             [ATGTCGGGCTGTTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
             using specific 5' and 3' primers and amplified by PCR. The PCR
             product was digested with SfiI and size selection was performed to
             exclude fragments <1.5kb.The SfiI-digested PCR product was cloned
             into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
             the DraIII sites can be used to isolate the cDNA insert. Libraries
             were constructed by Sugano et al.(University of Tokyo, Institute of
```

```
             Medical Science). Custom primer used for sequencing
             ( 5' end primer [CTTCTGCTCTAAAAGCTGCG];
             3' end primer [CGACCTGCAGCTGCACACA] ).
FEATURES             Location/Qualifiers
     source          1..2814
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
                     /clone="QccE-10361"
                     /sex="male"
                     /tissue_type="cerebellum cortex"
                     /clone_lib="macaque brain cDNA library QccE"
                     /dev_stage="adult"
     CDS             91..2502
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAB12255.1"
                     /db_xref="GI:9967105"
                     /translation="MGQPQHRAYPRQLFLLTAASPMAATTHRTLELMRWHRGTARCFS
                     FGLGPTCHQLQLGLSALSRGQAYFLRPGQRLQPMLVQALRKALEPALSDISVDWFVPD
                     TVEALLTPREIPALYPGDQLLGYCSLFRVDGFRPRPPGGQEPGWQSLGGGVFPSPEEA
                     PSAVSPGTEPTGTSEPLGTGTVSAELSSPWAAGDLERTGTDALIDPVTDPGPNPSSDT
                     AIWRRIFQSSYIREQYVLTHCSASPEPGPCSTSSESPSGSQGPSGGSSAPLEPPSQQ
                     GCRSLAWGEPAGSRSCPLPAPTPAPFKVGALSTEVLGQHRAALAGRSLSSPPGRANQ
                     VPGRPRKPSLGAILDGPSPEPGQOLGGQLDDSGSLLSPAPMDWDMLMEPPPLFTAVPP
                     SGEPAPPAVPPQAPRCHVVIRGLGEQPMCWEVGVGLETLWGPGDGSQOPPSPPVREAA
                     WDQALHRLTAASVVRDNEQLALRGGAETTADRGHARRCWLRALQTSKVSSAPSCFTCP
                     VAVDATTREVLPGVLQVCSSEPAEPPGTPPAAHSRLDAAPLPTVVYSKGAWDSDRNGN
                     SKCALGDAATPMEGPRCPPPRSPRSLGRRIKLCRPDLGQANNSEGIDHDYLPLVRL
                     QEAPGSFRLDAPFCAAVRISQERLCRASPFNVIRASLSPTSASLPWALLGPGVGQGDS
                     ATASCSPSPSGSSEGPGQVDSGRGSDTEASEGAEGLGGTDLRGRTWATAVALAWLEHR
                     CAAAFGEWELTAAKADCWLRAQHLPDGLDLAALKAAARGLFLLLRHWDQNLQLHLLCY
                     SPANV"
BASE COUNT     487 a    983 c    827 g    517 t
ORIGIN

Query Match            73.5%;  Score 1110.4;  DB 9;  Length 2814;
Best Local Similarity  93.6%;  Pred. No. 1.3e-190;
Matches 1190;  Conservative   0;  Mismatches  56;  Indels  26;  Gaps  2;
```

```
Qy    239 ggccatgcccgaggtgctggctccagcccttcagacaagtaaggtcagctctgcccc 298
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1537 GGCCATGCCCGAGGTGCTGGCTTCGAGCCCCTTCAGACACAAGTCAGCTCTGCCCCT 1596

Qy    299 tcctgcttcacttgccctgtagctggatgctactactagggaggtcctgcctggggcc 358
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1597 TCCTGCTTCACTTGCCCTGTAGCTGTGGATGCTACCACTAGGGAGGTCCTGCCTGGGGTC 1656

Qy    359 ctgcaggtgtgcagctcaagagcccgctgagccccaggaaccccctccctgcctctcacagc 418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1657 CTGCAGGTGTGCAGCTCAGAGCCAGCTGAGCCCCCAGGAACCCCCTCCCTGCCGCGTCACAGC 1716

Qy    419 catctagatgcagctcctcgccccactgtgtctactctaaaggactcagagaggcctct 478
          |||||||||||||||||||||||||||||||||||||||||||||||||-----------
Db   1717 CGTCTAGATGCAGCTCCTCTCGCCTACTGTTGTCTACTCTAA----------------- 1757

Qy    479 ccagcaggcgcctgggactcggaccaaaatgcaactccaagcgtgcttttggggggaccct 538
          ||||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db   1758 -----AGGCGCCTGGGACTCGGACAGAAAATGGCAACTCCAAGTGTGCTTTGGGGGACGGT 1812

Qy    539 gccaactcccacggaagtcctcgcccgccccactcccccgtcctccctgtcggctcagcatg 598
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1813 GCCACTCCCCATGGAAGGTCCTCGCTGCCCACCTCCCCCGTTCTCCCTCTCCGGCTTAGCCTG 1872

Qy    599 gggccgccgtcacaaactctgtagccccggcagccaggccaacaacagtgaaggcagc 658
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1873 GGGCCGGCCGTCACAAACCTCTGTAGACCTGACCTGGGCCAGGCCAGGCCAACAACAGTGAAGGCCATC 1932

Qy    659 gaccatgactacctgccttggtggctgcagaggcaccagctccttccgcctgac 718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1933 GACCATGACTACCTGCCCCTGGTGCGGCTGCGAGGAGGCACCAGGCTCCTTCCGCCCTGAC 1992

Qy    719 gcgcccttctgcgccgctgcgcatctcgagcggaggagcgcctcgcgtgcctctgcgcccttt 778
```

AUTHORS Loh,P., Qi,S., Ford,B., Hine,R. and Roe,B.A.
TITLE   Direct Submission
JOURNAL Submitted (06-FEB-2001) Department of Chemistry And Biochemistry,
         The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
         OK 73019, USA
COMMENT   On Jan 29, 2002 this sequence version replaced gi:15147174.
          -------------- Genome Center
          Center: Department Of Chemistry And Biochemistry
          The University of Oklahoma
          Center code:OOKNOR

          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1   2862: contig of 2862 bp in length
          *   2863   2962: gap of unknown length
          *   2963   7764: contig of 4802 bp in length
          *   7765   7864: gap of unknown length
          *   7865   34582: contig of 26718 bp in length
          *   34583   34682: gap of unknown length
          *   34683   59368: contig of 24686 bp in length
          *   59369   59468: gap of unknown length
          *   59469   95673: contig of 36205 bp in length
          *   95674   95773: gap of unknown length
          *   95774   143253: contig of 47480 bp in length
          *   143254   143353: gap of unknown length
          *   143354   198696: contig of 55343 bp in length.
FEATURES   Location/Qualifiers
     source   1..198696
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="rp23-101h1"
                     /clone_lib="rp23-101h1"
BASE COUNT   52616 a 48584 c 46801 g 50087 t   608 others
ORIGIN

Query Match   31.2%; Score 471.8; DB 2; Length 198696;
Best Local Similarity   76.1%; Pred. No. 1.1e-75;
Matches 655; Conservative 0; Mismatches 182; Indels 24; Gaps 5;

Qy   661 ccatgactacctgccctggtgcggctgcaggaggcaccaggctcctccgcctggacgc 720
           ||| ||| || ||||||||||||||||||||||||||||| || |||||||||||||| |||
Db 173106 CCACCCTTCTCTCCTCCTCCTCAGGTACGACTGCAAGAGGACCAGGCTCCTTCCGCCTGGATGA 173047

Qy   721 gcccttctggccgctgtggcgcatccgcaggaggcgcctctgccgtgccgcccttgc 780
           ||||| ||| |||||||||||||||||||||||||||| ||||||||||||||||||
Db 173046 ACCCTTCTGTGCTGCTGTGCATTCCTCAAGAACGCCTCAGACAGCCTCGCCCTTTGC 172987

Qy   781 cgtgcaccgcggcgcagcctcagcccaactcggcctcattgcctggcactcttcggccc 840
           || |||||| ||||||||||||||||||||||| ||||| || |||||||||||||||||
Db 172986 CGCACACCGTGCCAGCCTCAGCCCACCACCTAGCATCATCCTCCTTGGCATTTCTGAGCCC 172927

Qy   841 tggtgttggccaggtgacagtgccacgtcccagcccgtccccagctcgggctc 900
           |||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db 172926 TGGTATTGGCCAAGGTGACAGTGCCACAGCCTTCCTGCCACCAGTCCCCAGCTCCAGGTTC 172867

Qy   901 tgaggggccaggccaggtggacagtgggcgggggctcagacaccgaggcctccgatggggc 960
           ||||||||||||||| |||||||||| |||||||||||||||||||| |||||||||||||
Db 172866 TGAGGGTCCAGGCCAGGTAGCAGTGGCAGTGGGAGGGCCTCAGATACTGAGGCCTCAGAAGGAAT 172807

Qy   961 ggaagggctggcggcaccgacctgcggggccggcaggacctggccactgcctgagcactcgc 1020
           ||||||||||||||| ||||||||||||||||| ||||||||||||||||||||||| | |||
Db 172806 GGAAAGGCAGGACAGCTCTGATCTTGCTGTGGGCGCACTTGGCGCACAGCTGTGGCCCTGGC 172747

Qy   1021 ctggctggagcaccgatgccgcgctgcctcgacgagtgggaactgacagcggccaagc 1080
           ||||||||||||||| ||| ||||||||||| |||||||||||||||| |||||||||||

Db   1993 GCGCCCTTCTGTGCCGCTGTGCGGCATCTGCGGACGGAGGACGGCCTCTGCGCGCGGCCTCGGCCCTTT 2052
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Qy   779 gcgtgcaccgcgcagcctcagcccacctcggcctcattgcctggcacttctgggc 838
           ||||| || |||||||||||||||||||||||||||||||||||| ||| |||
Db   2053 GCCGTGCACCGTGCCAGCCTGCCAGCCTCAGCCCACCTCGGCCTCATTGCCTGGGCACTTCTAGGC 2112
Qy   839 cctggtgttggccaggtgacagtgccacggcctcccgagccccagtcgggc 898
           ||||||| | |||||||||||||||||||||||| |||||||||||||||
Db   2113 CCTGGTGTTGGTCAGGTGACAGTGCCACAGCCTCCTGCAGCCCGTCCCCAGCTCGGGC 2172
Qy   899 tctgaggggccaggccaggtggacagtggcgggggctcagacaccgaggcctccgatggg 958
           || ||||| || ||||||||||||||||| ||||||||| ||||||||| |||
Db   2173 TCCGAGGGTCCGGGCCCAGTGGACAGTGGGCGGGCGGGCTCAGATACCGAGGCCTCGGAGGGG 2232
Qy   959 gcggaagggctggggcgcaccgacctgcggggccgacctgggccactgccgtagcactc 1018
           |||||||||||||| ||||||||||||||||| ||||||||||||||||||||||| ||
Db   2233 GCGGAAGGGCTGGGGCGGCACCGACCTGCGGGCCGACCTGGGCCACTGCCGTGCGCTC 2292
Qy   1019 gcctgctggagcaccgatgacgcgcctgccctcgacgagtgggaactgacagcggccaag 1078
           ||||||||||||||| ||||| ||| ||||| ||||| |||||||||| ||||||||||
Db   2293 GCCTGGCTAGAGCACCGATGCGCTGCTGCCCTTGGCGAGTGGGAGTGGGAACTGACAGCAGCCAAG 2352
Qy   1079 gctgattgctggctgcggggccagcacttgctgctgacggccttgactggcgccctcaag 1138
           ||||||||||| ||||| |||||||||||| |||||| ||||||||||||||||||||||
Db   2353 GCTGATTGCTGGCTGCGGGGCCCAGCAGCTTGCCTGACGCACTTGCCTGACGGCCGCCCCTCAAG 2412
Qy   1139 gcccgcagcccgagggctcttcctgctactgcgccactgggaccaaaacctgcagctacac 1198
           |||||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   2413 GCCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCACTGGGACCAGAACCTGCAGCTACAC 2472
Qy   1199 ctgctgctacagccagcgaacgtgaaggctgccccctgctggccttggggctggcggcc 1258
           |||||||||||| ||||| ||||||||||| |||||||||||| ||||||||| ||||||
Db   2473 CTGCTGTGCTACAGCCCAGCAAACGTGTGAAGCTGCCCCGTGCTGCTGCTTGGGTTGGCGCC 2532
Qy   1259 ccaccaacactcaagtcactgcgcccagggctggcctctctggtgctggaaagtgt 1318
           ||||||||||||| ||||||||||||||||||||||||| |||||||||||||||||
Db   2533 CCACCCCAACACACT--AGACACTGCCGCCCCAGGACTGGCCTTCGGTGGTGGTCTGTGCTGGGAAGGTGT 2590
Qy   1319 aggctggtgccagcctgtccccactgctcttactccctcccggtggagccctcttgcccc 1378
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   2591 AGGCTCGGTGCCAGCCTGTCCCCACTGCTTCGTTCTACTCCCCTACTCCCTAGAGCCCTTGTGCCCC 2650
Qy   1379 cacaaaaagtcctgctgctgtctctcctcctctcccactcacactcccctcc 1438
           |||| |||| |||||||||| |||||||||||||| ||||||| ||||||||
Db   2651 CACAAGAAGTGCCTGCCGTCTGTCTCTGCTCTCTCCCCGCTCATCCCCCCACACACCTCCCCCTCC 2710
Qy   1439 atcctctgagctccctgcaacacagtggaaggtagagagccacagtccccaaatcctat 1498
           |||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2711 ATCCTCTGAGCTCCCTGCAACACAGTGGACAGTGGGAAGGGGAGAGCCACAGTCCCCAAATCCTAT 2770
Qy   1499 gcaataagtgc 1510
           ||||||||||
Db   2771 GCAATAAAGTGC 2782

RESULT 8
AC087898/c
LOCUS   AC087898   198696 bp   DNA   linear   HTG 29-JAN-2002
DEFINITION Mus musculus chromosome 16 clone rp23-101h1, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC087898
VERSION   AC087898.13 GI:18390230
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 198696)
AUTHORS   Loh,P., Qi,S., Ford,B., Hine,R. and Roe,B.A.
TITLE   Mus musculus Chromosome 16 BAC Clone rp23-101h1
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 198696)

```
Db 172746 GTGGCTGGAACACCGGCTGCGTGCAGCTGCCTTTGCGAGTGGGAACTGACAGCTTCCAAAGC 172687
          ||||||||||||||||| ||||| |||||||| ||| |||||||| ||||||| ||||||||
Qy   1081 tgattggctggctgcgggccagcacttgcctgacggcctggacggcggccctcaaggc    1140
          ||||||||||||||||| ||||| ||| ||| ||| |  ||||| ||| |||| |||| ||
Db 172686 TGATTGCTGGCTGCGAGAGCCCAACACCTGCCCAGACACGGACTTGACTGCTCTTAAAGC 172627
Qy   1141 cgcagccgagggctcttcctgctactggcactgggacaaaacctgcagctacacct    1200
          || ||||| |||||||| ||| ||||||| || ||||||| || |||||||||||| |
Db 172626 TGCTGCCCGGGGGCTCTTTTCTGTTACTGCGACACTGGGACCAGAATCTGCAGCTACACTT 172567
Qy   1201 gctgtgctacagcccagcgaacgtgtgaaggctgcccccctgcttgggctggcgcccc    1260
          ||||||| ||||||| ||| ||| ||| ||||||| || ||| |||||| |||| |||||
Db 172566 GCTGTGTTACAGCCCTTGCGAATGTGTAA---GGCTGCTTGCAGCTCTGGCTAGCGACCC 172511
Qy   1261 acccaacacactcaagtcactgccgcccagggctggcctcttggtgctgggaaagtgtag  1320
          |   |||||| || ||| |||||| ||||||  ||||| ||| || |||||||||| | ||
Db 172510 CCAACAC-----AGTCACTGCCACCCAGGCCAGTCTCTTGGTGTGCTGGGAAAGAATAA 172458
Qy   1321 gctggtgccagcc-------tgtccccactgcttcttactcctccctggagccc      1369
          | ||||| |||||        || |||||||| ||  |  |  || | || | ||
Db   2457 GTTGCTGTGTGCCGCCCTCCCATTGCTTCACATACACACACCACACCCCTCCCAACAT 172398
Qy   1370 tcttgccccacaaaaagtgcctgctgctctctcctcctccacccccactcaca       1429
          ||| | ||| ||||||||||||| |||| || || ||| | ||| || || | ||
Db 172397 TCTTGTCACCACAGAAAGTGCCTGCCTGTGCTCTCT-TCTCTCCTCCCCACTCCATACC 172339
Qy   1430 ctccctccatcctctgagctcctgcaacagtggaagggtagagagcacagtcccc      1489
          | |||||||| | || |||||| | ||||| || || || |||||||| ||| ||||
Db 172338 CCCTCCTATGTGCTGTCTGAGCTCTCTATAGGATAATGGAAGAGGAGAGCTACGGT-CCC 172280
Qy   1490 aatcctatgcaataaagtgc                                          1510
          |||||||||||||||||||||
Db 172279 AAATCCTATGCAATAAAGTGC 172259
```

RESULT   9
AC074027/c

```
LOCUS       AC074027        199359 bp    DNA    linear   HTG 25-JAN-2002
DEFINITION  Mus musculus chromosome 16 clone RP23-113H11 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION   AC074027
VERSION     AC074027.7  GI:12313746
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 199359)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       High Throughput Mouse Sequencing
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 199359)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUL-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12043571.
            -----------Genome Center
            Center: Harvard Partners Genome Center
            Center Code: HPGC
            Web site:
            http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous
            e.html
            Contact: gnktm@capecod.bwh.harvard.edu
            ------------Summary Statistics
            Center project name: ADH
            Sequencing vector: pUC18; L08752
            Chemistry: Dye-terminator Big Dye; 100%
```

```
*Consensus quality:   184639 at least Q20
*Consensus quality:   178829 at least Q30
*Consensus quality:   168661 at least Q40
 Estimated insert size:     agarose-FP - N/A
**Estimated insert size: 198699 - sum-of-contigs
 Quality coverage:     agarose-FP - N/A
 Quality coverage:     4.8 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1    12682 contig of 12682 bp in length
*  12683    12702 gap of unknown length
*  12703    23292 contig of 10590 bp in length
*  23293    23312 gap of unknown length
*  23313    39092 contig of 15780 bp in length
*  39093    39112 gap of unknown length
*  39113    51108 contig of 11996 bp in length
*  51109    51128 gap of unknown length
*  51129    60218 contig of 9090 bp in length
*  60219    60238 gap of unknown length
*  60239    70060 contig of 9822 bp in length
*  70061    70080 gap of unknown length
*  70081    79260 contig of 9180 bp in length
*  79261    79280 gap of unknown length
*  79281    90181 contig of 10901 bp in length
*  90182    90201 gap of unknown length
*  90202    98036 contig of 7835 bp in length
*  98037    98056 gap of unknown length
*  98057   105849 contig of 7793 bp in length
* 105850   105869 gap of unknown length
* 105870   114552 contig of 8683 bp in length
* 114553   114572 gap of unknown length
* 114573   123701 contig of 9129 bp in length
* 123702   123721 gap of unknown length
* 123722   131069 contig of 7348 bp in length
* 131070   131089 gap of unknown length
* 131090   138295 contig of 7206 bp in length
* 138296   138315 gap of unknown length
* 138316   144002 contig of 5687 bp in length
* 144003   144022 gap of unknown length
* 144023   149683 contig of 5661 bp in length
* 149684   149703 gap of unknown length
* 149704   154301 contig of 4598 bp in length
* 154302   154321 gap of unknown length
* 154322   158504 contig of 4183 bp in length
* 158505   158524 gap of unknown length
* 158525   161217 contig of 2693 bp in length
* 161218   161237 gap of unknown length
* 161238   164558 contig of 3321 bp in length
* 164559   164578 gap of unknown length
* 164579   167331 contig of 2753 bp in length
* 167332   167351 gap of unknown length
* 167352   171444 contig of 4093 bp in length
* 171445   171464 gap of unknown length
* 171465   175319 contig of 3855 bp in length
* 175320   175339 gap of unknown length
* 175340   178350 contig of 3011 bp in length
* 178351   178370 gap of unknown length
* 178371   181271 contig of 2901 bp in length
* 181272   181291 gap of unknown length
* 181292   183851 contig of 2560 bp in length
* 183852   183871 gap of unknown length
* 183872   187525 contig of 3654 bp in length
* 187526   187545 gap of unknown length
* 187546   189662 contig of 2117 bp in length
* 189663   189682 gap of unknown length
* 189683   190563 contig of 881 bp in length
```

```
*    190564  190583: gap of unknown length
*    190584  192520: contig of 1937 bp in length
*    192521  192540: gap of unknown length
*    192541  194141: contig of 1601 bp in length
*    194142  194161: gap of unknown length
*    194162  196034: contig of 1873 bp in length
*    196035  196054: gap of unknown length
*    196055  198800: contig of 2746 bp in length
*    198801  198820: gap of unknown length
*    198821  199359: contig of 539 bp in length.
```

```
FEATURES          Location/Qualifiers
source            1..199359
                  /organism="Mus musculus"
                  /strain="C57BL6/J"
                  /db_xref="taxon:10090"
                  /chromosome="16"
                  /clone="RP23-113H11"
                  /sex="male"
misc_feature      1..12682
misc_feature      12703..23292
                  /note="assembly_name:Contig91"
misc_feature      23313..39092
                  /note="assembly_name:Contig90"
misc_feature      39113..51108
                  /note="assembly_name:Contig89"
misc_feature      51129..60218
                  /note="assembly_name:Contig88"
misc_feature      60239..70060
                  /note="assembly_name:Contig87"
misc_feature      70081..79260
                  /note="assembly_name:Contig86"
misc_feature      79281..90181
                  /note="assembly_name:Contig85"
misc_feature      90202..98036
                  /note="assembly_name:Contig84"
misc_feature      98057..105849
                  /note="assembly_name:Contig83"
misc_feature      105870..114552
                  /note="assembly_name:Contig82"
misc_feature      114573..123701
                  /note="assembly_name:Contig81"
misc_feature      123722..131069
                  /note="assembly_name:Contig80"
misc_feature      131090..138295
                  /note="assembly_name:Contig79"
misc_feature      138316..144002
                  /note="assembly_name:Contig78"
misc_feature      144023..149683
                  /note="assembly_name:Contig77"
misc_feature      149704..154301
                  /note="assembly_name:Contig76"
misc_feature      154322..158504
                  /note="assembly_name:Contig75"
misc_feature      158525..161217
                  /note="assembly_name:Contig74"
misc_feature      161238..164558
                  /note="assembly_name:Contig73"
                  clone_end:T7
misc_feature      164579..167331
                  vector_side:left"
                  /note="assembly_name:Contig71
misc_feature      167352..171444
                  clone_end:SP6
                  vector_side:right"
misc_feature      171465..175319
                  /note="assembly_name:Contig70"
misc_feature      175340..178350
                  /note="assembly_name:Contig69"
misc_feature      178371..181271
                  /note="assembly_name:Contig68"
misc_feature      181292..183851
                  /note="assembly_name:Contig67"
```

```
misc_feature      183872..187525
                  /note="assembly_name:Contig66"
misc_feature      187546..189662
                  /note="assembly_name:Contig65"
misc_feature      189683..190563
                  /note="assembly_name:Contig64"
misc_feature      190584..192520
                  /note="assembly_name:Contig63"
misc_feature      192541..194141
                  /note="assembly_name:Contig62"
misc_feature      194162..196034
                  /note="assembly_name:Contig61"
misc_feature      196055..198800
                  /note="assembly_name:Contig60"
misc_feature      198821..199359
                  /note="assembly_name:Contig59"
                  /note="assembly_name:Contig58"
```

```
BASE COUNT    52621 a   47137 c   46866 g   51719 t    1016 others

Query Match           31.2%;    Score 471.8;   DB 2;   Length 199359;
Best Local Similarity  76.1%;   Pred No. 1.1e-75;
Matches  655;   Conservative  0;   Mismatches 182;   Indels  24;   Gaps  5;

Qy   661 ccatgactacctgccttgtgcggctgcaggagcaccaggctccttccgcctggacgc  720
             ||  |||  |  || || |||||||||||||||||||||||||||||||||||  |
Db 71624 CCACCTCTTCTCTCCTCCAGTACGACTGCAAGAGCACCAGGCTCCTTCCGCCTGGATGA 71565

Qy   721 gcccttcgccgcctgtgccatctgcagagagcgcctctgccgtgcctcgccctttgc  780
             | |||||||||||||||||| |||||||| ||| ||||||||||||||||||||
Db 71564 ACCCTTCGTGCTGCTGTGTGCATTCCTCAAGAACGCCTGTGTGCAGCCTCGCCCTTTGC 71505

Qy   781 cgtgcacgcgccagctcagcccaactcggcctcattgcctggcgacttctgggccc  840
          || ||||| |||||||| |||||| ||||| ||||||||||||| ||||| ||||| |||
Db 71504 CGCACACCGTGCCAGCCTCAGCCCCACCTCAGCCATCATCTCCTTGGGCATTTCTGAGCCC 71445

Qy   841 tggtgttggccaggtgacagtggcgggggctcagacaccgagcctccgatggggc  900
          |||| |||||| |  || |||||||  ||||  || || |||| || |||  ||| ||
Db 71444 TGGTATTGGCCAAGGTGACAGTGCCACAGTGGGAGGGGCTCAGACTACTGAGGGCTCAGGTC 71385

Qy   901 tgaggggccaggccaggtggacagtggcggggctcagagtggagctggaactggaagc  960
          ||||| |||||||||||| |||||||||  ||||||||| ||||  | |||||  ||
Db 71384 TGAGGGTCCAGGCCAGGTAGACAGTGGGAGGGGCTCAGACTACTGAGGGCTCAGAAGGAAT 71325

Qy   961 ggaagggctgggggcaccgaccgggcggggccgacctgggcactgccgtagcactcgc  1020
          ||||| |||||||  ||||| ||||| |||||| |  | || ||||||| |||| ||
Db 71324 GGAAAGGCAGGAGCAGGAGCAGCTCTGATCTTCGTGTGGGGCCACTTGGGCCACAGCTGTGGCCCTGGC 71265

Qy  1021 ctggctggagcaccgatgcgccgctgccttcgacgagctgacgacgcggccaaggc  1080
          ||||| |||||| | || ||||  | |||| ||| |||||| || |||||| ||||
Db 71264 GTGGCTGGAACACCGCTGCTGCCTTTGGCGAGCTGCCTGGAGTGGGAACTGACGACGCTTCCAAAGC 71205

Qy  1081 tgattgctggctgcgggcccagcacttgcctgacggcctgacctggccgccctcaaggc  1140
          || |||||||||||| | ||| ||  |||||||||||| |||||||||| ||| |||||
Db 71204 TGATTGCTGGCTGGCTGGAGACCCAAACCTGCCAGCCCCAACCTGCAGGCCGACTTGATCTCATCGTCTTAAAGC 71145

Qy  1141 cgcagccgaggggctcttcctgctactggcaaaactgcagctacacct  1200
          || ||| |||||||||| |||| || |||||| |||| |||||||||||
Db 71144 TGCTGCCCGGGGGCTCTTTCTGGTTACTGCGACACTGGGACCACAGAATCTGCAGCTACACTT 71085

Qy  1201 gctgtgctacagccagcgaacgtgtgaaggctgtccccctgcttggctggctggcgccccc  1260
          |||||| ||||||||||  | |||  |||||||| ||  | |||||||| |||||||| ||||
Db 71084 GCTGTTGTTACAGCCCTTCAGCCCTTCGAATGTGTAA---GGCTGCTTGCAGCTCTTGGCTAGGCACCC 71029

Qy  1261 accaacacactcaagtcactgccgccagggctggcctcttggtgctgggaaagtgtag  1320
          ||  || ||  |||| |  || |||  |  | ||| |  | |||||| |||||  |
Db 71028 CCAACAC------AGTCACTGCCACCCGGGCCACCAGTCTCTTGGCTGCTTGCGTGTGGGAAAGAATAA 70976

Qy  1321 gctggtgccagc--------tgtccccactgcttcttactccctccctggagccc  1369
          ||||| ||||||        |||| || |||| ||||| | ||||| || |||| |
Db 70975 GTTGGTGTCTGCGTCCGCTCCATTGCTTCACATACACACACACCCCTCCAACAT 70916

Qy  1370 tcttgccccacaaaagtgcctgctgctctcctcctctctcccaccactcaca  1429
          || ||| |||| |   |||||| ||||| |||||| || || | ||  |
```

```
Db 70915 TCTTGTCACCACAGAAAGTGCCTGCCTGTCTCT-TCTCTCTCCTCCCACTCCATACC 70857
              |||||||||||||||| ||  |||||| |  ||||||||||| |||| || |||
Qy  1430 ctccccccatctctgagctcctcaacacagtgaaggtagagagccacagtcccc 1489
              | ||| | |||||||||| | |||||||| | |||||||| ||| || |||
Db 70856 CCCTCCTATGTGCTCTGAGCTCTCTATAGGATAATGGAAGAGGAGAGCTACGGT-CCC 70798

Qy  1490 aaatcctatgcaataagtgc 1510
              |||||||||||||| ||||||
Db 70797 AAATCCTATGCAATAAAGTGC 70777
```

RESULT 10
AC073708

LOCUS   AC073708   219564 bp   DNA   linear   HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-187G15, WORKING DRAFT SEQUENCE, 28
            unordered pieces.
ACCESSION   AC073708
VERSION   AC073708.1 GI:8810325
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 219564)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 219564)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 1808839
            Center clone name: RPCI-23_187G15

            Summary Statistics
            Consensus quality: 191425 bases at least Q40
            Consensus quality: 206405 bases at least Q30
            Consensus quality: 209449 bases at least Q20
            Estimated insert size: 204000; agarose-fp estimation
            Estimated insert size: 216864; sum-of-contigs estimation
            Quality coverage: 7.33 in Q20 bases; agarose-fp estimation
            Quality coverage: 6.9 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 28 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1   1579: contig of 1579 bp in length
          *   1580   1679: gap of unknown length
          *   1680   2767: contig of 1088 bp in length
          *   2768   2867: gap of unknown length
          *   2868   3993: contig of 1126 bp in length
          *   3994   4093: gap of unknown length
          *   4094   5096: contig of 1003 bp in length
          *   5097   5196: gap of unknown length
          *   5197   6332: contig of 1136 bp in length
          *   6333   6432: gap of unknown length
          *   6433   7435: contig of 1003 bp in length
          *   7436   7535: gap of unknown length
          *   7536   8686: contig of 1151 bp in length
          *   8687   8786: gap of unknown length
          *   8787   9904: contig of 1118 bp in length
          *   9905 10004: gap of unknown length
          *   10005 11906: contig of 1902 bp in length
          *   11907 12006: gap of unknown length
          *   12007 14261: contig of 2255 bp in length
          *   14262 14361: gap of unknown length
          *   14362 16859: contig of 2498 bp in length
          *   16860 16959: gap of unknown length
          *   16960 19396: contig of 2437 bp in length
          *   19397 19496: gap of unknown length
          *   19497 22518: contig of 3023 bp in length
          *   22520 22619: gap of unknown length
          *   22620 24754: contig of 2135 bp in length
          *   24755 24854: gap of unknown length
          *   24855 30882: contig of 6028 bp in length
          *   30883 30982: gap of unknown length
          *   30983 35339: contig of 4357 bp in length
          *   35340 35439: gap of unknown length
          *   35440 41737: contig of 6298 bp in length
          *   41738 41837: gap of unknown length
          *   41838 45907: contig of 4070 bp in length
          *   45908 46007: gap of unknown length
          *   46008 54514: contig of 8507 bp in length
          *   54515 54614: gap of unknown length
          *   54615 60249: contig of 5635 bp in length
          *   60250 60349: gap of unknown length
          *   60350 69789: contig of 9440 bp in length
          *   69790 69889: gap of unknown length
          *   69890 79032: contig of 9143 bp in length
          *   79033 79132: gap of unknown length
          *   79133 91048: contig of 11916 bp in length
          *   91049 91148: gap of unknown length
          *   91149 102627: contig of 11479 bp in length
          *   102628 102727: gap of unknown length
          *   102728 116363: contig of 13636 bp in length
          *   116364 116463: gap of unknown length
          *   116464 131945: contig of 15482 bp in length
          *   131946 132045: gap of unknown length
          *   132046 155515: contig of 23470 bp in length
          *   155516 155615: gap of unknown length
          *   155616 219564: contig of 63949 bp in length.
FEATURES   Location/Qualifiers
     source   1..219564
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-187G15"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT   55924 a 52052 c 51899 g 56983 t   2706 others
ORIGIN

Query Match   31.2%; Score 471.8; DB 2; Length 219564;
Best Local Similarity   76.1%; Pred. No. 1.1e-75;
Matches 655; Conservative   0; Mismatches 182; Indels   24; Gaps   5;

Qy   661 ccatgactacctgccttgtgcggctgcaggaggaggaggcaggcacaccgccctccgcctggacgc 720
                 || || |||||||||||||| ||||||| ||||||| |||||||||||||||||||||
Db   190117 CCACCTTCTGTCTCTCTCTCAGGTACGACTGCAAGAGGCACCAGGCTCCTTCCGCCTGGATGA 190176

Qy   721 gccttctcgccgccgtgtgcgcatctcgcaggaggaggcgcctctgccgtgctgcctgccctttgc 780
                 | ||||||| || | ||| | || | | ||||||| || |||||||||||||||||
Db   190177 ACCCTTCTGTGTCGTGCTGTGTGCATTCCTCAAGAACGCCGTGTGTCGAGCCCGCCCTTTGC 190236

Qy   781 cgtgcacgcgcgcagcctcagcccaccccggcctcattgccctgggcacttctgggccc 840
                || |||||||||| | |||||||||||||| ||| |||||| ||||||||||||
Db   190237 CGCACACCGTGCCAACCGTGCCAGCCCCACCTCAGCATCATCCTTGGGCATTCTGAGCCC 190296

Qy   841 tggtgttggccagggtgacagtgccacggccacggccctcccagctcccagtcgggctc 900
                |||| |||||| ||| ||||||| || | ||||||||||| ||||||||||| || ||
Db   190297 TGGTATTGGCCAAGTGACAGTGACAGTGCCACAGTGCCACAGCCAGCTCAGGTTC 190356

Qy   901 tgaggggcaggcaggtggacaatgggcgggggctcagacaccgaggggctccgatgggc 960
                ||||||| ||||| ||| ||| |||| || |||| |||||||| |||||||| ||| |
Db   190357 TGAGGGTCCAGGCCAGTAGACAGTGGGAGGAGGGGGCTCAGAATACTGAGGCCTCAGAGGAAT 190416
```

```
Qy    961 ggaagggctgggcggcaccgacctgcggggccggacctgggccactgccgtagcactcgc 1020
          || ||| ||||| || | | || || || ||| ||||| || | ||||| |||||
Db 190417 GGAAAGGCAGGACAGCTTCTGATCTTCGTGGGCGCACTTGGGCCACAGCTGTGGCCCTGGC 190476

Qy   1021 ctggctggagcaccgatgcgcgcgtccttcgacgagtgggaactgacagcggccaaggc 1080
          || ||||||| ||||| |  |||||| || ||||||||||||||| ||  ||| |||
Db 190477 GTGGCTGGAACACCGGCTGTGCAGCTGCCTTTGGCGAGTGGGAACTGACAGCTTCCAAAGC 190536

Qy   1081 tgattgctggctgcgggccagcacttgcctgacggcctgacctggccgccctcaaggc 1140
          ||||||||||| ||| ||| || ||| ||| | |||| |  || || ||| || |||
Db 190537 TGATTGCTGGCTGGCTGAGAGCCCAACAACCTGCCAGACGGACTTGATCTGACTGCTCTTAAAGC 190596

Qy   1141 cgcagcccgagggctcttcctgctactgcccactgggaccaaaacctgcagctacacct 1200
          ||| ||||| |||  ||| |  |||||||| ||||| ||| ||| || ||||||||||| ||
Db 190597 TGCTGCCCGGGGGGCTCTTTCTGTTACTGCGACACTCGGACCCAGAATCTGCAGCTACACTT 190656

Qy   1201 gctgtgctacagccagcgaacgtgtgaaggctgacccctgctgtggctggctgcgccccc 1260
          | ||| ||||| | ||||   |||| ||   |  |  | ||| |||||| || ||| |||
Db 190657 GCTGTGTTACAGCCCTTCGAATGTGTAA----GGCTGCTTGCAGCTCTGGCTAGCGACCC 190712

Qy   1261 acccaacacactcaagtcactgcgcgcccagggctggcctcttggtctgctgggaaagtgtag 1320
          ||||||||| |||||| ||||||||| || ||||||||||||| ||
Db 190713 CCAACAC------AGTCACTGCCACCCCAGGGCCAGTCTCTTGGTGCTGGGAAAGAATAA 190765

Qy   1321 gctgggtgccagcc----------tgtccccactgcttcttactccctccctggagccc 1369
          ||||||| |  |||||                 || | || ||  |  ||  || ||
Db 190766 GTTGGTGTCTGCGCTCCGCCTCCATTGCTTCACATACACACACACCCCTCCCAACAT 190825

Qy   1370 tcttgccccacaaaagtgctgcctgtgctctctcctcctccccaccccactcaca 1429
          ||  | ||||    ||| |||||||||||||||||||||   ||| | | ||| ||
Db 190826 TCTTGTCACCCAGAAAGTGCCTGCCTGTGCTCTCT-TCTCTCTCTCCCACTCCCATACC 190884

Qy   1430 ctccccccatcctctgagctccctgcaacacagtggaagggtagagagccacagtcccc 1489
          || | ||| | |||||||||  || |  | ||||||| ||||  |||  ||| | ||| |||
Db 190885 CCCTCCTATGTGCCTCTGAGCTCTGAGCTCTATAGGATAATGGAAGAGGAGAGCTACGGT-CCC 190943

Qy   1490 aaatcctatgcaataaagtgc 1510
          || |||||||||||||||||||
Db 190944 AAATCCTATGCAATAAAGTGC 190964

Search completed: September 19, 2002, 04:23:18
Job time: 12968 sec
```

```
              GenCore version 4.5
      Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        September 19, 2002, 01:06:10 ; Search time 235.79 Seconds
                                             (without alignments)
                                   11002.408 Million cell updates/sec

Title:         US-09-786-136-4
Perfect score: 1511
Sequence:      1 tccagagtgctgaatttctg........atccatgcaataagtgca 1511

ng table:      IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:    3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     N Geneseq_032802.7
               1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
               9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
              10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
              11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
              12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
              13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
              14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
              15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
              16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
              17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
              18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
              19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
              20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
              21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
              22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
              23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
              24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

G-te hits 2-4

                          SUMMARIES

Result No.		Score	%Query Match	Length	DB	ID	Description
c	1	1511	100.0	1511	21	AAA08405	Human neurotransmi
c	2	1217.8	80.6	2777	22	AAK84116	Human immune/haema
c	3	604.8	40.0	864	22	AAI94103	Human neuroblastom
c	4	370.8	24.5	882	22	AAI94102	Human neuroblastom
c	5	60.8	4.0	10732	21	AAA10594	Gene encoding a su
c	6	60.6	4.0	1337	21	AAZ17263	Human gene express
c	7	58.2	3.9	114955	20	AAX53491	Human adenosine A1
c	8	55.4	3.7	114955	20	AAX53491	Human adenosine A1
		53	3.5	1127	21	AAA02477	Human colon cancer
c	10	50	3.3	33718	22	AAK65257	Human immune/haema
c	11	50	3.3	33718	22	AAK86411	Human immune/haema
	12	49.8	3.3	1995	22	AAF63462	Oligonucleotide #2
	13	49.8	3.3	4686	22	AAK80905	Human immune/haema
	14	49.8	3.3	4403765	22	AAI99683	Mycobacterium tube
	15	49.8	3.2	4411529	22	AAI99682	Mycobacterium tube
	16	48.2	3.2	4684	22	AAK80907	Human immune/haema
c	17	47.2	3.1	2016	23	AAS69866	DNA encoding novel
	18	46	3.0	15872	18	AAT68715	Streptomyces venez
	19	46	3.0	15872	21	AAZ87283	S. venezuelae vep
c	20	45.6	3.0	1218	21	AAA02488	Human colon cancer
c	21	45.4	3.0	4403765	22	AAI99683	Mycobacterium tube
c	22	45.4	3.0	989	21	AAA02539	Human colon cancer
	23	45	3.0	109519	22	AAS08693	Micromonospora DNA
	24	45	3.0	125401	22	AAD17186	Streptomyces nours
c	25	44.4	3.0	1214	22	AAF60302	Human zalpha48 CDN
	26	44.4	2.9	58857	21	AAA58471	Nucleotide sequenc
	27	44.2	2.9	11220	21	AAZ87298	S. venezuelae macr
	28	44.2	2.9	36778	21	AAZ87318	S. venezuelae pik
	29	44.2	2.9	37948	21	AAZ87285	S. venezuelae pik
	30	44.2	2.9	38506	21	AAA75633	Human polynucleoti
	31	44.2	2.9	38506	21	AAA56001	Nucleotide sequenc
c	32	44	2.9	1140	21	AAA96228	cDNA encoding a ma
	33	44	2.9	2064	14	AAQ52638	Streptomyces fradi
	34	43.6	2.9	3600	17	AAT13227	Thermostable enzym
	35	43.4	2.9	1359	20	AAZ17254	Human gene express
	36	43.4	2.9	1860	14	AAQ50632	Brain factor-2. R
	37	43.2	2.9	1000	21	AAA02484	Human colon cancer
c	38	43.2	2.9	1352	22	AAI58126	Human polynucleoti
c	39	43.2	2.9	1776	15	AAQ86111	Human alpha-1A adr
c	40	43.2	2.9	1776	16	AAT11601	Human alpha-1A adr
c	41	43	2.8	1474	15	AAQ70311	Mouse NF-ATp. Mus
c	42	43	2.8	1531	21	AAA59261	DNA encoding a hum
	43	43	2.8	2998	21	AAF21180	Human low adenosin
	44	43	2.8	2998	21	AAA35058	Human adenosine re
	45	43	2.8	3478	17	AAT36866	Human transcriptio

                              ALIGNMENTS

```
RESULT 1
AAA08405
ID   AAA08405 standard; cDNA; 1511 BP.
XX
AC   AAA08405;
XX
DT   13-JUL-2000  (first entry)
XX
DE   Human neurotransmitter associated nucleotide sequence SEQ ID NO:4.
XX
KW   Human; neurotransmitter; diagnosis; antiparkinsonian; neuroleptic;
KW   anticonvulsant; nootropic; tranquillizer; neuroprotective; cytostatic;
KW   antidepressant; antidiabetic; gynaecological; immunosuppressive;
KW   neurological disorder; hormone-related disease; Parkinson's disease;
KW   schizophrenia; epilepsy; female reproductive disorder; ss.
KW   attention deficit disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             596..1228
FT                   /*tag= a
FT                   /product= "neurotransmitter associated protein"
XX
PN   WO200012685-A2.
XX
PD   09-MAR-2000.
XX
PF   26-AUG-1999;  99WO-US19615.
XX
PR   01-SEP-1998;  98US-0144952.
PR   01-SEP-1998;  98US-0155194.
```

XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Walker MG, Volkmuth W, Klingler TM;
XX
DR   WPI; 2000-375619/32.
DR   P-PSDB: AAY82530.
XX
PT   A new purified polynucleotide comprising a gene that is coexpressed
PT   with neurotransmitter-processing-specific genes in biological samples
PT   for diagnosing, treating or preventing neurological and hormone-related
PT   diseases -
XX
PS   Claim 2: Page 34-35; 36pp; English.
XX
CC   The present invention describes a purified polynucleotide comprising a
CC   gene that is coexpressed with one or more neurotransmitter-processing-
CC   specific genes in biological samples. The neurotransmitter-processing-
CC   specific genes are L-tyrosine hydroxylase (TH), aromatic amino acid
CC   decarboxylase (AADC), dopamine beta-hydroxylase (DBH), nicotinic
CC   acetylcholine receptor alpha3 subunit precursor (nAChR-alpha3),
CC   secretogranin I and II, Rab3a, human cocaine and amphetamine regulated
CC   transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), and ARIX
CC   homeodomain protein. The present sequence represents a neurotransmitter
CC   associated nucleotide sequence from the present invention. The
CC   polynucleotides from the present invention can have antiparkinsonian,
CC   neuroleptic, anticonvulsant, nootropic, tranquilizer, neuroprotective,
CC   cytostatic, antidepressant, antidiabetic, gynaecological and
CC   immunosuppressive activities. They can be used for diagnosing, treating,
CC   preventing or evaluating therapies for neurological and hormone-related
CC   diseases, particularly Parkinson's disease, schizophrenia, epilepsy,
CC   female reproductive disorders and attention deficit disorder. The gene
CC   products are therapeutic proteins and targets of therapeutics against
CC   the diseases.
XX
SQ   Sequence 1511 BP; 252 A; 529 C; 435 G; 295 T; 0 other;

Query Match   100.0%; Score 1511; DB 21; Length 1511;
Best Local Similarity   100.0%;
Matches 1511; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 tccagagtgctgaatttctggcagccagggggtctcgctctgctctgctatggttgaagatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 tccagagtgctgaatttctggcagccagggggtctcgctctgctctgctatggttgaagatt 60

Qy   61 cagcctgatcactccttgtcaaagtggggagtcgagctggttctggcctagcaaggc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 cagcctgatcactccttgtcaaagtggggagtcgagctggttctggcctagcaaggc 120

Qy   121 aatctactctgttaaagtcgatagagggagaagctgggagaacactgcaattttccta 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 aatctactctgttaaagtcgatagagggagaagctgggagaacactgcaattttccta 180

Qy   181 tgagccctctgtagagggtggtggtgtggggtgtggggctccacagactgctcctggttccacagg 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 tgagccctctgtagagggtggtggtgtggggtgtggggctccacagactgctcctggttccacagg 240

Qy   241 ccatgccggaggtgctggcttcagccccttcagacaagtaaggtcagctcagccctc 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ccatgccggaggtgctggcttcagccccttcagacaagtaaggtcagctcagccctc 300

Qy   301 ctgctcacttgcctgtagctgtgatgctactactagggaggtcctgcctgggggccct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ctgctcacttgcctgtagctgtgatgctactactagggaggtcctgcctgggggccct 360

Qy   361 gcaggtgtgcagctcagagcccgtgagcccccaggaacccctcctgcctctcacagcca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 gcaggtgtgcagctcagagcccgtgagcccccaggaacccctcctgcctctcacagcca 420

Qy   421 tctagatgcagctcctctgcccactgttgtctactctaaaggacttcagagaggctctcc 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 tctagatgcagctcctctgcccactgttgtctactctaaaggacttcagagaggctctcc 480

Qy   481 agcaggcggctcggggactcggaccaaaatggcaactccaagcgtctttggggggaccctgc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 agcaggcggctcggggactcggaccaaaatggcaactccaagcgtctttggggggaccctgc 540

Qy   541 cactccacggaagtcctcgccgccccactcctccctgtcgcctccccgtctcggctcagcatggg 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 cactcccacggaagtcctcgccgccccactcctccctgtcgcctccccgtctcggctcagcatggg 600

Qy   601 ccgcgtcacaaactctgtagccctgaccgggcggccaacaacagtgaaggcagcga 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ccgcgtcacaaactctgtagccctgaccgggcggccaacaacagtgaaggcagcga 660

Qy   661 ccatgactacctgccctgtgcggctgcaggaggcaccagcgtcctccgcctggacgc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ccatgactacctgccctgtgcggctgcaggaggcaccagcgtcctccgcctggacgc 720

Qy   721 gccctctgcgccgctgcgcatctgcaggagcgcctcgccgtgcctgccctcggcccttgc 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 gccctctgcgccgctgcgcatctgcaggagcgcctcgccgtgcctgccctcggcccttgc 780

Qy   781 cgtgcacgcgccagcctcagcccaccctcggcctcattgccctggcgactctgggccc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 cgtgcacgcgccagcctcagcccaccctcggcctcattgccctggcgactctgggccc 840

Qy   841 tggtgttggccaggtgacagtgcacggcctcctgcagccgtccccccagctcggggctc 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 tggtgttggccaggtgacagtgcacggcctcctgcagccgtccccccagctcggggctc 900

Qy   901 tgaggggccaggccagtggacagtggggctcggggagccaggcctccgatggggc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 tgaggggccaggccagtggacagtggggctcggggagccaggcctccgatggggc 960

Qy   961 ggaagggctgggcggcaccagcatggcgccgctgcctttcgacgagtggaactgacagcggccaaggc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ggaagggctgggcggcaccagcatggcgccgctgcctttcgacgagtggaactgacagcggccaaggc 1020

Qy   1021 ctggctggagcacccagatgcgccgctgcctttcgccgagtggaactgacagcggccaaggc 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ctggctggagcacccagatgcgccgctgcctttcgccgagtggaactgacagcggccaaggc 1080

Qy   1081 tgattgctgctggcggcggcccagcacttgcctgacctgacctgctgacctgaccatgcctggccgccctcaaggc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tgattgctgctggcggcggcccagcacttgcctgacctgacctgctgacctgaccatgcctggccgccctcaaggc 1140

Qy   1141 cgcagcccggaggctctttctgctactgcgcactggaccaaaacctgcagctacacct 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 cgcagcccggaggctctttctgctactgcgcactggaccaaaacctgcagctacacct 1200

Qy   1201 gctgtgctacagcccagcgaacgtgtgaaggctgaaggctgccccctgctgctgggctggcgcccc 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gctgtgctacagcccagcgaacgtgtgaaggctgaaggctgccccctgctgctgggctggcgcccc 1260

Qy   1261 acccacacactcaagtcactgcgccccaggtcggccagggctggcctcttggtgtctgggaaagtgtag 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 acccacacactcaagtcactgcgccccaggtcggccagggctggcctcttggtgtctgggaaagtgtag 1320

Qy   1321 gctggtgccagcctgctctcttactcctcctggagccccttgccccca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gctggtgccagcctgctctcttactcctcctggagccccttgccccca 1380

Qy   1381 caaaaagtgcctgcctgtctctcctcccctctcccctccaccccactcacactccctccat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 caaaaagtgcctgcctgtctctcctcccctctcccctccaccccactcacactccctccat 1440

Qy   1441 cctctggcctccctgcaacagtggaaggtagagagcacagtcccaaatcctatgc 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 cctctgagctcctgcaacagtggaaggtagagagcacagtcccaaatcctatgc 1500

Qy   1501 aataagtgca 1511
          ||||||||||
Db   1501 aataagtgca 1511

```
RESULT   2
AAK84146/c
ID   AAK84146 standard; DNA; 2777 BP.
XX
AC   AAK84146;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38958.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
     Homo sapiens.
     WO200157182-A2.
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.        PR   12-SEP-2000; 2000US-0231968.
PR   04-FEB-2000; 2000US-0180628.        PR   14-SEP-2000; 2000US-0232397.
PR   24-FEB-2000; 2000US-0184664.        PR   14-SEP-2000; 2000US-0232398.
PR   02-MAR-2000; 2000US-0186350.        PR   14-SEP-2000; 2000US-0232399.
PR   16-MAR-2000; 2000US-0189874.        PR   14-SEP-2000; 2000US-0232400.
PR   17-MAR-2000; 2000US-0190076.        PR   14-SEP-2000; 2000US-0232401.
PR   18-APR-2000; 2000US-0198123.        PR   14-SEP-2000; 2000US-0233063.
PR   19-MAY-2000; 2000US-0205515.        PR   14-SEP-2000; 2000US-0233064.
PR   07-JUN-2000; 2000US-0209467.        PR   14-SEP-2000; 2000US-0233065.
PR   28-JUN-2000; 2000US-0214886.        PR   21-SEP-2000; 2000US-0234223.
PR   30-JUN-2000; 2000US-0215135.        PR   21-SEP-2000; 2000US-0234274.
PR   07-JUL-2000; 2000US-0216647.        PR   25-SEP-2000; 2000US-0234997.
PR   07-JUL-2000; 2000US-0216880.        PR   25-SEP-2000; 2000US-0234998.
PR   11-JUL-2000; 2000US-0217487.        PR   26-SEP-2000; 2000US-0235484.
PR   14-JUL-2000; 2000US-0217496.        PR   27-SEP-2000; 2000US-0235834.
PR   14-JUL-2000; 2000US-0218290.        PR   27-SEP-2000; 2000US-0235836.
PR   26-JUL-2000; 2000US-0220963.        PR   29-SEP-2000; 2000US-0236327.
PR   26-JUL-2000; 2000US-0220964.        PR   29-SEP-2000; 2000US-0236367.
PR   14-AUG-2000; 2000US-0224518.        PR   29-SEP-2000; 2000US-0236368.
PR   14-AUG-2000; 2000US-0224519.        PR   29-SEP-2000; 2000US-0236369.
PR   14-AUG-2000; 2000US-0225213.        PR   29-SEP-2000; 2000US-0236370.
PR   14-AUG-2000; 2000US-0225214.        PR   02-OCT-2000; 2000US-0236802.
PR   14-AUG-2000; 2000US-0225266.        PR   02-OCT-2000; 2000US-0237037.
PR   14-AUG-2000; 2000US-0225267.        PR   02-OCT-2000; 2000US-0237038.
PR   14-AUG-2000; 2000US-0225268.        PR   02-OCT-2000; 2000US-0237039.
PR   14-AUG-2000; 2000US-0225270.        PR   02-OCT-2000; 2000US-0237040.
PR   14-AUG-2000; 2000US-0225447.        PR   13-OCT-2000; 2000US-0239935.
PR   14-AUG-2000; 2000US-0225757.        PR   13-OCT-2000; 2000US-0239937.
PR   14-AUG-2000; 2000US-0225758.        PR   20-OCT-2000; 2000US-0240960.
PR   14-AUG-2000; 2000US-0225759.        PR   20-OCT-2000; 2000US-0241221.
PR   18-AUG-2000; 2000US-0226279.        PR   20-OCT-2000; 2000US-0241785.
PR   22-AUG-2000; 2000US-0226681.        PR   20-OCT-2000; 2000US-0241786.
PR   22-AUG-2000; 2000US-0226868.        PR   20-OCT-2000; 2000US-0241787.
PR   22-AUG-2000; 2000US-0227182.        PR   20-OCT-2000; 2000US-0241808.
PR   23-AUG-2000; 2000US-0227009.        PR   20-OCT-2000; 2000US-0241809.
PR   30-AUG-2000; 2000US-0228924.        PR   20-OCT-2000; 2000US-0241826.
PR   01-SEP-2000; 2000US-0229287.        PR   01-NOV-2000; 2000US-0244617.
PR   01-SEP-2000; 2000US-0229343.        PR   08-NOV-2000; 2000US-0246474.
PR   01-SEP-2000; 2000US-0229344.        PR   08-NOV-2000; 2000US-0246475.
PR   01-SEP-2000; 2000US-0229345.        PR   08-NOV-2000; 2000US-0246476.
PR   05-SEP-2000; 2000US-0229509.        PR   08-NOV-2000; 2000US-0246477.
PR   05-SEP-2000; 2000US-0229513.        PR   08-NOV-2000; 2000US-0246478.
PR   05-SEP-2000; 2000US-0230437.        PR   08-NOV-2000; 2000US-0246523.
PR   06-SEP-2000; 2000US-0230438.        PR   08-NOV-2000; 2000US-0246524.
PR   08-SEP-2000; 2000US-0231242.        PR   09-NOV-2000; 2000US-0246525.
     -SEP-2000; 2000US-0231243.          PR   08-NOV-2000; 2000US-0246526.
     -SEP-2000; 2000US-0231244.          PR   08-NOV-2000; 2000US-0246527.
          2000; 2000US-0231413.          PR   08-NOV-2000; 2000US-0246528.
          2000; 2000US-0231414.          PR   08-NOV-2000; 2000US-0246532.
             0; 2000US-0232080.          PR   08-NOV-2000; 2000US-0246609.
              ; 2000US-0232081.          PR   08-NOV-2000; 2000US-0246610.
                                         PR   08-NOV-2000; 2000US-0246611.
                                         PR   08-NOV-2000; 2000US-0246613.
                                         PR   17-NOV-2000; 2000US-0249207.
                                         PR   17-NOV-2000; 2000US-0249208.
                                         PR   17-NOV-2000; 2000US-0249209.
                                         PR   17-NOV-2000; 2000US-0249210.
                                         PR   17-NOV-2000; 2000US-0249211.
                                         PR   17-NOV-2000; 2000US-0249212.
                                         PR   17-NOV-2000; 2000US-0249213.
                                         PR   17-NOV-2000; 2000US-0249214.
                                         PR   17-NOV-2000; 2000US-0249215.
                                         PR   17-NOV-2000; 2000US-0249216.
                                         PR   17-NOV-2000; 2000US-0249217.
                                         PR   17-NOV-2000; 2000US-0249218.
                                         PR   17-NOV-2000; 2000US-0249244.
                                         PR   17-NOV-2000; 2000US-0249245.
                                         PR   17-NOV-2000; 2000US-0249264.
                                         PR   17-NOV-2000; 2000US-0249265.
                                         PR   17-NOV-2000; 2000US-0249297.
                                         PR   17-NOV-2000; 2000US-0249299.
                                         PR   17-NOV-2000; 2000US-0249300.
                                         PR   01-DEC-2000; 2000US-0250160.
```

PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI: 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
     useful for preventing, diagnosing and/or treating cancers and
     metastasis -
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 2777 BP; 569 A; 789 C; 861 G; 558 T; 0 other;

Disclosure; SEQ ID NO 38958; 3071pp + Sequence Listing; English.

Query Match   80.6%; Score 1217.8; DB 22; Length 2777;
Best Local Similarity   85.2%;
Matches 1508; Conservative   0; Mismatches   2; Indels 259; Gaps 3;

Qy   1 tccagagtgctgaattctggcagcagaggggctcttgtctgctatggggttgaagatt   60
          ||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db   1778 TCCAGAGTGCTGAATTTCTGGCAGCAGAGGGGCTCTTGTCTGCTATGGGTTGAAGATT 1719

Qy   61 cagcctgatcacttcttgtcaaagtggggagtctgagctggttctgggcctagcaagc 120
          ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db   1718 CAGCCTGATCATCACTTCTTGTCAAAGTGGGGAGTCTGAGCTGGTTCTGGGCCTAGCAAGGC 1659

Qy   121 aatctactctgttaagtcgatagaggagaagctgggaggagaacactgcaatttttcta 180
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |
Db   1658 AATCTACTCTGTTAAAGTCGATAGAGGAGAAGCTGGGAGGAGAACACTGCAATTTTTTCTA 1599

Qy   181 tgagccctctgtagagggtggtgtgtggggctccacagactgctcctggttccacagg 240
          |||||||||||||||||||||| ||| |||||||||||||||||||||| |||||||||
Db   1598 TGAGCCCTCTGTAGAGGGTGGGTGGTGGGGCTCGGGGCTCCACAGACTGCTCCTTGGTTCCACAGG 1539

Qy   241 ccatgccggaggtgctggcttcagcccttcagacaagtaaggtcagctctgcccccctc 300
          ||||||| ||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db   1538 CCATGCCCGGAGGTGCTGGCTTCGGGCTTCAGACAGTAAGGTCAGCTCTGCCCCCTC 1479

Qy   301 ctgtcttcacttgccctgtagctgtggatgctactactactagggaggtcctgcctgggggccct 360
          || ||||||||||||||||||||||||||||| |||||| |||||||||||||||||||||||
Db   1478 CTGCTTCACTTGCCCTGTAGCTGTGGATGCTGTGGATGCTACTAGGGAGGTCCTGCCTGGGGGCCCT 1419

Qy   361 gcaggtgtgcagct--   374
          |||||||||||||||
Db   1418 GCAGGTGTGCAGCTTCAGGTAGGGGCCTTCTTCTTGTGTACCTCTCAGGTGCAGCTCTCACCT 1359

Qy   375 ------------cagagccggct   385
                      |||||||||||
Db   1358 CACTGGCCACTCTCCTGGACAAACCTGACTCCTGTCCCATGTGCCCCTGCGAGAGCCCGCT 1299

Qy   386 gagccccaggaacccctcctgctctcacagccatctagatgcagctcctctgcccact 445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1298 GAGCCCCAGGAACCCCTCCTGCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACT 1239

Qy   446 gttgtctactctaa-------   459
          ||||||||||||||
Db   1238 GTTGTCTACTCTAAAGGTAACACCCAAAGGTAGGGAAAGGTAGGGGCACTTGGGCTTAG 1179

Qy   460 -------------aggacttcagagaggctctccagcagg   486
                       ||||||||||||||||||||||||||
Db   1178 AGACCGCCCCTGCACTGATGATCCCCACTGCAGGACTTCAGAGAGGCTCTCCAGCAGG 1119

Qy   487 cgcctggagtcggaccaaaatgcaactccaagcgtgtctttgggggaccctgccactcc 546
          ||||||| |||||||||||||||||||||||||||| ||||||| ||||||||||||||
Db   1118 CGCCTGGGACTCGGACCAAAATGCAACTCCAAGCGTGCTTTGGGGACCCTGCCACTCC 1059

Qy   547 cacggaagtcctcgccgccacctcccgtcctccctgtcggctcagcatgggcgccgcg 606
          ||||||||||||||||||||||||||||||||| |||| |||||||||||||||||||||
Db   1058 CACGGAAGTCCTCGCCGCCACCTCCCGTCCTCCCGTCCGGCTCAGCATGGGCGCCGCG 999

Qy   607 tcacaaactcgtagccccgaccgggccaggccaacaacagtgaagcagcagcgaccatga 666
          ||||||||||||||||| |||| |||||||||||||||||||||| |||||||||||||
Db   998 TCACAAACTCGTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGA 939

Qy   667 ctacctgccctt------------   678
          ||||||||||||
Db   938 CTACCTGCCCTTGGTGAGGACTCCGGAGGTGGTGCGCCGCGGGGCGGGGCGCTG 879

Qy   679 -------------ggtgcggctgcaggaggaggaccag   701
                       |||||||||||||||||||||||||
Db   878 TTTCAGCTCGGCTTCCTCCCCACCTCCTCTCCCCCAGGTGCGGCTGCAGGAGGAGGACCACCAG 819

Qy   702 gctccttcgcctggacggccccttcgcgccgctggcgatcctgcggcatctccgaggagcgcctct 761
          |||||||||||||||||||||||||| |||||||| | |||| || ||||||||||||||||||||
Db   818 GCTCCTTCGCCTGGACGGCCCCTTCTCGCGCTGGGTGTCGCATCCGCAGGAGGCGCCTCT 759

Qy   762 gccgtggcctcgccctttcgccgtgcacgcgcagcctcagccccaactcggcctcattgc 821
          |||||| |||||||||| ||||||||||||||||||||||||||| ||||||||||||||
Db   758 GCCGTGCCTCGCCCTTTGCCGTGCACGCGCAGCCTCAGCCCCACCTCGGCCTCATTGC 699

Qy   822 cctgggcacttctggcctggtgttggccaggtgacagtgcacggcctcctgcagcc 881
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   698 CCTGGGCACTTCTGGCCTGGTGTTGGCCAGGTGACAGTGCACGGCGTCCTGCAGCC 639

Qy   882 cgtccccagctcgggctctgagggggtctgagggggctgagggcgtggggggctcagaca 941
          ||||||||||||||| |||||||| ||||| ||||| |||| |||||||| |||||| |||
Db   638 CGTCCCCAGCTCGGGCTCTGAGGGGCTCAGGGGCTCAGGGCTGGACAGTGGACAGACA 579

Qy   942 ccgaggcctcgatgggcggagaggctggccggcaccgaccctcgggggccggaccctggg 1001
          |||||||||| |||||| ||||||| ||||| ||||||||||| ||||||||| |||||||
Db   578 CCGAGGCCTCCGAGGGCCGGAAGGGCGTGGGCGGCCACCGACCTCGGGGGCCCGGAACCTGGG 519

Qy   1002 ccactgcgtagcactcgcctggagcactcgcctggagagcaccgatcgccgctcgagtggg 1061
          |||||| ||||||||||||||||||| ||| |||||| |||||||||||| |||||||||||
Db   518 CCACTGCCGTAGCACTCGCCTGGAGCAGTCGCAGCACCGATCGCCCGCCTCGCCGCAGTGGG 459

Qy   1062 aactgacagcggccaaggctgattgctgctgcgggccagacacttgcctgacggccttg 1121
          |||||||||||||||||| ||||||||| ||||||||||| ||||||||||||||||||
Db   458 AACTGACAGCGGCCAAGGCCAGAGGCTGATTGCTGGCGGGCCCAGCACTTGCCTGACGGCCTTG 399

Qy   1122 acctggccgccctcaaggccgcagcccgagggcttcctgctactgcgccactgggacc 1181
          |||||||||||||||||||| ||||||||||| |||||||||||||||||||||||||
Db   398 ACCTGGCCGCCCTCAAGGCCCCGAGGGCTCTTCTGCTACTGCGCCACTGGGACC 339

Matches 668; Conservative 0; Mismatches 85; Indels 3; Gaps 2;

```
Qy    756 gcctctgccgtgcctcgcccttgccgtgcaccggcaccggcagccgcagcccactcggcct 815
          ||||||||||||||||||||     |||| | ||| ||| |||||||||||  | |-| ||
Db    803 GCCCTTTTCCGTCCTTNGCCTTNCCGTGCAACGGGGC--AGCTTAGCCCAACTTGGCCTT 746

Qy    816 cattgccctggcacttctggccctggttggcccaggtgacagtgccac ggcc-tcc 874
          |||| ||  ||||||||||||| ||||||||||||||||   |-||| |||
Db    745 NATNGCCTTGGAATTTTTGGCCCTGTTTTTGCCCAGGTGACAGTTCCAAGCCTTCT 686

Qy    875 tgcagccgtccccagctgggctcgggctctgagggcgcaggccagcaagtgggacgagtgggcggggc 934
          |||||||||| | ||| | ||-||||  |||| ||| ||| |||| ||| | |||||
Db    685 TGAAGCCCGTCCCCAAGTTNGGGNTTGAGGGNCCAGGCCAGTTGGACATTGGGCGGGGT 626

Qy    935 tcagacaccgagggctccgatgggcggaaaggctgggcggcaccgacctgcgggggccgg 994
          |||||||||||||   ||| ||| ||  ||| |  | | ||| | |||||| |||||
Db    625 TCAGACACGGAGCCTTCGGAGGGGGCGANGGGTTGGGCGGAACGGACTTGCGGGCCCGG 566

Qy    995 acctgggcactgccgtagcactcgcctggctggagcaccggcaccggccgctgccttcgac 1054
          || ||-|| | ||| | || |||| |||||||| || |||||| || || ||| | ||
Db    565 ACTTGGNCAATTGCNGTAGCATTNGCTTGGTTGGAGCACCGATGCGCNGATGCNTTNGAG 506

Qy   1055 gagtggaactgacagcggccaaggctgattgctggctgcgggccagcacttgcctgac 1114
          |||||||| ||||||| |||||||||||||||| |||||||||||||||||||||||||
Db    505 GAGTGGGAATTGACACGCGCCAAGGCTGATTGTTGGTTGCGGCCCAGCACTTGCCTGAC 446

Qy   1115 ggcctgacctggcgccctcaagccccagccgcgaggctctcctgctactggccac 1174
          ||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db    445 GGCNTTGACTTGGCCGCCNTCAAGGCCCAGCCGCCGAGGGCTCTTCTTGCTACTGCGCCAC 386

Qy   1175 tgggaccaaaacctgcagctacacctgtgctacacggccagcgaacgtgtgaaggctg 1234
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    385 TGGGACCAAAACCTGCAGNTACACCTGCTGTCGTACAGCCCAGCGAACGTGTGAAGGCTG 326

Qy   1235 cccctgtgcttgggctgcgcgccccacccaacacactcaagtcactgccgcccagggct 1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 CCCCCTGTCGTTGGGCTTGGGCTGGCGCGCCCCACCCACCACTCAAGTCACTGCCGCCCAGGGCT 266

Qy   1295 ggcctcttggtgtctgggaaagtgtaggctggtgtgccagcctgtcccccactgcttcttact 1354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 GGCCTCTTGGTGCTGGGAAAGTGTAGGCTGGTGCTGGCCAGCCTGTCCCCCACTGCTTCTTACT 206

Qy   1355 ccctccctggagccctcttgccccacaaaaagtgcctgcctgctctctcctcctcct 1414
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    205 CCCTCCCTAGAGCCCTTCTTGCCCCACAAAAAGTGCTGCTGTCGTGTGTCTCTCCCTCTCCT 146

Qy   1415 cccaccccactcacactccctccatcctctgagctccctgacacagtggaagggtag 1474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 CCCACCCCACTCACACTTCACACTCCCTCCATCCTCTGAGCTCCCTGCAACACAGTGGAAGGGTAG 86

Qy   1475 agagcacagtccccaaatcctatgcaataaagtgc 1510
          |||||||||||||||||||||||||||||||||||
Db     85 AGAGCCACAGTCCCCAAATCCTATGCAATAAAGTGC 50
```

RESULT 4
AAI94102
ID   AAI94102 standard; cDNA; 882 BP.
XX
AC   AAI94102;
XX
DT   13-NOV-2001 (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 177.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166719-A1.
XX
PD   13-SEP-2001.
XX

```
Qy   1182 aaaacctgcagctacacctgctgtgctacagccagcgaacgtgtgaaggctgcccctg 1241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    338 AAAACCTGCAGCTACACCTGCTGTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCTG 279

Qy   1242 ctgcttgggctggcgcgccaccaacacactcaagtcactgccgcccagggctggcctct 1301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 CTGCTTGGGCTGGCGCGCCCCACCAACACACTCAAGTCACTCACTGCCGCCCAGGGCTGGCCTCT 219

Qy   1302 tggtgctgggaaagtgtaggctggtgcagcgctgtccccactgcttcttactccctcc 1361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 TGGTGCTGGGAAAGTGTAGGCTGGTGCGCCAGCCTGTCCCCACTGCTTCTTACTCCCTCCC 159

Qy   1362 tggagccctcttgccccacaaaaagtgcctgcctgtctctcccctcctccccaccc 1421
          ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 TAGAGCCCTTCTTGCCCCACAAAAAGTGCCTGCTGTCCTGTCTCTCCCCTCCTCCCCACCC 99

Qy   1422 cactcacactcccctccatccctcgagctccctgcaacacagtgaagggtagagagcca 1481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 CACTCACACTCCCCTCCATCCCTCTGCAGCTCCCTGCAACACAGTGCAACACAGTGGAGAGCCA 39

Qy   1482 cagtccccaaatcctatgcaataaagtgc 1510
          ||||||||||||||||||||||||||||||
Db     38 CAGTCCCCAAAATCCTATGCAATAAAGTGC 10
```

RESULT 3
AAI94103/C
ID   AAI94103 standard; cDNA; 864 BP.
XX
AC   AAI94103;
XX
DT   13-NOV-2001 (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 178.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166719-A1.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-JP01629.
XX
PF   07-MAR-2000; 2000JP-0159195.
XX
PI   (CHIB-) CHIBA PREFECTURE.
     (HISM) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human
PT   neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT   for anti-cancer agents -
XX
PS   Claim 1; Page 168-169; 2979pp; Japanese.
XX
CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes.
XX

Sequence 864 BP; 188 A; 235 C; 257 G; 145 T; 39 other;

tch   40.0%;   Score 604.8; DB 22; Length 864;
    Similarity 88.4%;   Pred. No. 2.5e-125;
```

```
PF   02-MAR-2001; 2001WO-JP01629.
XX
PR   07-MAR-2000; 2000JP-0159195.
XX
PA   (CHIB-) CHIBA PREFECTURE.
PA   (HISM ) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human
PT   neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT   for anti-cancer agents -
XX
PS   Claim 1; Page 168; 2979pp; Japanese.
XX
CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes.
XX
SQ   Sequence 882 BP; 193 A; 222 C; 218 G; 210 T; 39 other;

Query Match             24.5%;  Score 370.8;  DB 22;  Length 882;
Best Local Similarity   98.9%;  Pred. No. 3e-73;
Matches   372;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 tccagagtgctgaattctgggcagccagggggctcttgctctgctatgggttgaagatt   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254 tccagagtgctgaattctgggcagccagggggctcttgctctgctatgggttgaagatt  313

Qy    61 cagcctgatcacttcctgtcaaagttggggagtctgagctgttctggccctagcaaggc  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 cagcctgatcacttcctgtcaaagttggggagtctgagctgttctgggcctagcaaggc  373

Qy   121 aatctactctgttaaagtcgatagagggagaagctgggagaacactgcaatttttcta  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 aatctactctgttaaagtcgatagagggagaagctgggagaacactgcaatttttcta  433

Qy   181 tgagccctctgtagaggggtgggtggtggggggtcccacagactgctccttggttccacagg  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 tgagccctctgtagaggggtgggtggtggggggtcccacagactgctccttggttccacagg  493

Qy   241 ccatgcccggaggtgctgcttcagcccttcagacaagtaaggtcagctctgcccccctc  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   494 ccatgcccggaggtgctgcttcagcccttcagacaagtaaggtcagctctgcccccctc  553

Qy   301 ctgcttcacttgccctgtagctgtggatgctactactaggagggtcctgcctgggcccct  360
         ||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   554 ctgcttcacttgccctgtanctggatgctactactactanggaagtcctgcctgggcccct  613

Qy   361 gcagtgtgcagctca  376
         ||| |||||||||||
Db   614 gcaagtgtgcagctca  629

RESULT 5
AAA10594/c
ID   AAA10594 standard; DNA; 10732 BP.
XX
AC   AAA10594;
XX
DT   29-JUN-2000 (first entry)
XX
DE   Gene encoding a subunit of cellulose synthase.
XX
KW   Cellulose synthase; cellulose production; increase yield; ds.
XX
OS   Vigna angularis.
XX
PN   JP2000060568-A.
XX
PD   29-FEB-2000.
XX
PF   26-AUG-1998; 98JP-0239998.
XX
PR   26-AUG-1998; 98JP-0239998.
XX
PA   (MIZU/) MIZUNO K.
PA   (OJIP ) OJI PAPER CO.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; AAY85179.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match             4.0%;   Score 60.8;   DB 21;  Length 10732;
Best Local Similarity   12.8%;  Pred. No. 0.00051;
Matches   141;  Conservative  433;  Mismatches  523;  Indels    7;  Gaps    2;

Qy   341 gaggtcctgcctgggccctgcagtgtgcagctcagacgccgctgagccccaggaaac  400
         ::|               :  ::||:   :  :  |::: |  ::::::|:
Db 10398 SRGNSTYSNSTSRNSTCCCYSDTBSRNSTSDTTTTSRCTTCYDASDNCCSTNSTYSNSTS 10339

Qy   401 cctcctgcctcacagccatctagatgcagctccctcgccccactgttgtcactctaaa  460
Db 10338 RNCNSTSRCNSTNCRCNSTYSCSRDYSNCCCYTYSSTYDASRCYSCNCNCNSTAKYSCC 10279

Qy   461 ggacttcagagaggtctccagcaggcgcctgggactcggaccaaaatggcaactccaag  520
Db 10278 YTSTYDATTYSNCSRCYTCTBCSRNCYSCTTYSSRSTSRSTYDADCSRCNSTYRAYRACR 10219

Qy   521 cgtgctttggggaccctgccactcccacggaaggtcctcgcgccccacctcccgtcct  580
Db 10218 CNSTNCSTSDRCSRGYSYDAYDATBNCYSYSNSTCCYTYSDCYTCSRCTBSDYDACNST 10159

Qy   581 ccctgtcggctcagcatgggccgcgtcacaaactctgtagccctgaccgggccaggc-  639
Db 10158 CCRCSTYSCYTYDACSTYASTDNCSRSRFTSTCYTAKYSYSTBSTNCYSDSRGSDNSTAK 10099

Qy   640 caacaacagtgaaggcagcgaccatgactacctgccttggtcgggctgcaggaggcacc  699
Db 10098 YASTTTYDAYDARCTTDNSTRCYRASTYDASRRCYSSTYRASRGTBSDDCCTBSRYSCYS 10039

Qy   700 aggctccttccgcctgacgcgccccttctgcgcgcgctgcgcatctcgcagagcgcct  759
Db 10038 YDASRYANCDTBCYTTBAKYRARCYDAYAKRCNSTSRYRAYSDSTCYTRCRCSRNCYSTY 9979

Qy   760 ctgcgtgcctgcgcccttgccgtgcaccgcagccctcagccccacctcggcctcatt  819
Db  9978 SYSSTYRASTTBTTYCYTCBCSRRCYSRYSSTCNCYSYCCYTSRYSSTTTNSTCRCTT 9919

Qy   820 cctcctgggcacttctgggccgccgtgttggccaggtgacagtgcacagccctctgcag  879
Db  9918 YSYNSTBTBYSYSSTYSSRGYSSSRGSDSRGNCYNSTNCYDASTSDTBYSRCCYTYSYS 9859

Qy   880 ccgtccccagctcgggggtctgaggggccagccagcccaggtggacagtgggcggggctcaga  939
```

Db   9858 STDSTDSTDSTNSTTBSDCYTTTBSRSTSDSTSTYRCSRSDYDATBSDNSTNCCYDASR 9799
                 :|:::|:|::::|::|::::::::::::|::::::::::::|::
Qy   940 caccgaggcctccgatggggcggaagggctgggcggcaccgacctgcggggcggacctg 999
                 :|:::|:|::::|::|:::::::::::::|::::::::::::|:
Db   9798 TBTBSTNCYARCYTBYDARCSRDSTYSSRGYDANSTSRYSSRYSSTYSSDSTYSAKYCAK 9739

Qy   1000 g------gccactgcgtagcactcgcctgctgagcaccgatgcgccgctgccttcga   1053
                 :::|::||:::::|:|:::::|::::::|::|:::::|::
Db   9738 STTBTBCYDAYDACYDAYDANCYSSDSTYTBYCSRRCCCYDAYSCSRYDARCYDACYS 9679

Qy   1054 cgagtgggaactgacagcggcaagctgccaaggctgattgctgcgcggcccagcacttgcctga 1113
                 ::|::|:|::   ::|:: |:||::|::||:|:::: :
Db   9678 YSNSTCYDATBTTSRYSTTYSYSNCYDATTSRCTBYSTBBTTSRCAKCTBDSTSTAKN 9619

Qy   114 cggcctgacctggccgccctcaagccgcagcccgaggctcttcctgctactgcgcca 1173
                 :::||::|:::::|::: ::|:|:||::::::|::::||:::|||::::|:
Db   618 STSYSTTRCTBYSSRSRGYSYCSRSRSRCYSCYTDSDSTCYSTYSTTYYAYSCTTSRGY 9559

Qy   1174 ctgggaccaaaacctgcagctacacctgctgctacagcccagcgaacgtgtgaaggct 1233
                 ::|||:::|: ||::| :|::||:|:|: :::
Db   9558 SYDASRSTSYSRCTTSYSYSSTDYSDCSDYSTTTBNSTYSSDSDCTBYSSDRCSRSDS 9499

Qy   1234 gcccctgctggctggcggcgcccaccaacactcaagtcactgccgcccaggc 1293
                 |:::||:::|:| ::|:|:::|::|:: :::|:
Db   9498 TCNCYSCSDSRYSTYDACYTYDAAKTBCYTYSDNCCNSTSRCYTNSTYSCSRTBSRNST 9439

Qy   1294 tggcctcttggtgctgggaaagtgtaggctggtgctgagcctgtccccactgcttcttac 1353
                 :|::| :::|::::::|::|:||::|::::|:
Db   9438 CCCTBTTSRGNCCYDAYDANSTSRYDAYYDACYYSYDASTTBYSYSCTBYSNSTYDAYSS 9379

Qy   1354 tcctccctggagccctcttgccccacaaaaagtgcctgctgtgctctcctcctcc 1413
                 ::|:|::|:|::: ::|: :|::|::|::|:|:|:||:::||:
Db   9378 RYSCYTCYTCDYSSTCYTYRCAKCTBCNSTSRAKSRTTTTYRCYCYTSRCYTBCYTCY 9319

Qy   1414 tccacccacctcacactccccctc 1437
                 |:|::||::|||:::::::|:|
Db   9318 TNCSRGCCCSRTTSRSRYDAYSCNC 9295

RESULT 6
AAZ17263
ID   AAZ17263 standard; cDNA; 1337 BP.
XX
AC   AAZ17263;
XX
     12-OCT-1999   (first entry)

Human gene expression product cDNA sequence SEQ ID NO:4735.

KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999; 99WO-US01619.
XX
PR   03-APR-1998; 98US-0080666.
PR   28-JAN-1998; 98US-0072910.
PR   24-FEB-1998; 98US-0075954.
PR   31-MAR-1998; 98US-0080114.
PR   03-APR-1998; 98US-0080515.
XX
PA   (CHIR) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI   Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI   Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI   Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI   Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR   WPI: 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 2250-2251; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;

Query Match   4.0%; Score 60.6; DB 20; Length 1337;
Best Local Similarity   22.0%; Pred. No. 0.00037;
Matches 236; Conservative 0; Mismatches 831; Indels 4; Gaps 1;

Qy   390 cccaggaaccctcctgcctctcacagccatctagatgcagctccctctgccactgttg 449
                | || || | | | || |||| ||| | | |||||||||||
Db   243 cnncnnnnnncnnnnnnnnnnnnncnnnnnnnnnnnnnnnnncnnnnnnnnncnnnnnnnn 302

Qy   450 tctactctaaaggacttcagagaggctctccagcaggcgcctgggactcggaccaaaatg 509
                | | | || || | | || || |||| || | || || | |||||| | |
Db   303 nnnnnnnnnnnnnnnnnncnnnnnnncnnnnnnnnnncncnnccccnncncnnnnnn 362

Qy   510 gcaactccaagtgcttgggggaccctgccactcccacggaagtcctcgcgccccac 569
                | | || || ||| ||| | || | | ||| || | | | | || ||
Db   363 ncnnnnncnncnnnncnccnnnnccncnnnnnnncncnnnnnnncccccccnncnnc 422

Qy   570 ctccccgtcctcctgtcggctcagcagcatgggccggccgtcacaaactctgtagcctgacc 629
                ||||| | | | || | | ||| |||| | | | ||| | | | | ||||||
Db   423 ncccnnncncccnnnnncncnnnnncncnnnnnncnnnnnncccccccnnnnnnnnnncncc 482

Qy   630 cgggcaggcaacacagtgaaggcagcgaccatgactacctgccttggtgcggctgc 689
                | | || || || | || ||| || | | ||| | |||||| || | ||| |
Db   483 ccnncncccccccccccncccccncccccccccnncccccccccnncccnnccccccnc 542

Qy   690 aggaggaccaggctccttcgcctgacgcgccttctgcgcgctgtgcgcatctgc 749
                || || || |||| || |||| | | | |||| || ||| || || | |||
Db   543 ncncncccccnnncncnnnnncccnnnnnnncccnnnccccccccnnncncnccccc 602

Qy   750 aggagcgc----tctgccgtgcctcgcctttgccgtgcaccgcgcagcctcagcccc 805
                || || | |||| || | || ||| | ||| | | || || | |||||
Db   603 cnccccccccnncccccccncncnccccccccccccccnnnnccccccccccccccc 662

Qy   806 acctcggcctcattgcctggcggcacttctgggcctggtggttggccagggtgacagtgcc 865
                 || | || | | ||||| || | | | ||| || || | ||||||| | |||||
Db   663 cccnnncncnccccccccccccccnnnccnccccccagctcggggctcggggcttcgaggggcctcgagggggccaggccagatggacact 722

Qy   866 acggccctgcagcccgtccccagctcgggctcggggtctgaggggccaggcaggtggacagt 925
                | | || | | || | ||| | | ||| | | |||

```
Db    723 ccncccccccnnccccnccccnnncccncnccccnnccccncnccccnncccnc 782
          |           |
Qy    926 gggcgggctcagacaccgaggcctccgatgggcggaaggctggcggcaccgacctg 985
          |        ||  ||
Db    783 cncncnnnnccccncccncnnnnnncnccccncnccncnccccnccncccccn 842
          |    |  ||
Qy    986 cggggccgacctgggccactgccgctagcactcgccgctggtggagcaccgatgcgccgct 1045
          |        |  |||  ||
Db    843 ccnnnnncncnccccnnncnccnnncnccccnccccnnccccnnncnccnnccncnn 902
          |    ||  |  |   |
Qy   1046 gccttcgacgagtggaactgacagcggccaaggctgattgctggctgcgggccagcac 1105
          |  |||  || |  ||   ||  |
Db    903 nnccccncccnnncnnncnnncnnncnnncnnncnncnnnnnncncnccnnnn 962
          | |||  |
Qy   1106 ttgcctgacggccttgacctgcgccgcctcaaggccgcagcccgagggctcttcctgcta 1165
          ||  ||   ||  | |  |  ||||  |
Db    963 cncncnnnnncncnnnnnnncccnnnccnnncnccnncnccnnnnnncnnnnnnccnn 1022
          |  |  |  |  ||  ||
Qy   1166 ctgccacctggacacaaaacctgcagctacacctgctgtgctacagccagcgaacgtg 1225
          | |  ||  |||  ||
Db   1023 cnncnnnnnnnccccncncnccccnncncnnnnnnctcnccccncncncncnnn 1082
          | |       |
Db   1226 tgaaggctgccccctgctgctggctggcgccccaccacactcaagtcactgccg 1285
          ||| | ||   ||  |  ||
Db   1083 cnncnncnnnnnncnnncncncnccccnnncncncnnnnnnncnnncnnnnnnnncn 1142
          | |  |   ||
Qy   1286 cccaggggctggcctcttggtgctggggaaagtagtaggctggtgccagcctgtccccactg 1345
          | | | | ||
Db   1143 cncncncnnnnnncncncncnnnnncncncncncncncnnnnncncnnnnnnccnc 1202
          |  |  |
Qy   1346 cttcttactccctccctgagccctcttgccccacaaaaagtgctgctgtgtctct 1405
          |    | |||||
Db   1203 nncncncncncnnnnncncncnccccccccccnnnnncnnnnncnnnnncncncncncnn 1262
          | |  |  |||||
Qy   1406 ccctctcctcccccacccccactcacactccctccctcctctcgagctccctgc 1456
          |   |||  |||  |   |  ||   |
Db   1263 ncncncnnnncccncccncncccnnnncnnnccccnnnnncncncccncc 1313
          | |        ||
```

```
RESULT 7
AAX53491/c
ID    AAX53491 standard; DNA; 114955 BP.
XX
AC    AAX53491;
XX
DT    05-JUL-1999 (first entry)
XX
DE    Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW    Antisense oligonucleotide; multiple target; antisense treatment;
KW    impaired respiration; inflammation; lung disease;
KW    pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW    acute asthma; allergy; asthma; impeded respiration;
KW    respiratory distress syndrome; pain; cystic fibrosis;
KW    pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW    chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW    colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW    hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW    prostate cancer; ss.
XX
OS    Synthetic.
XX
PN    WO9913886-A1.
XX
PD    25-MAR-1999.
XX
PF    17-SEP-1998;    98WO-US19419.
XX
PR    09-JUN-1998;    98US-0093972.
PR    17-SEP-1997;    97US-0059160.
XX
PA    (UYEC-) UNIV EAST CAROLINA.
XX
```

```
PI    Nyce JW;
XX
DR    WPI; 1999-229400/19.
XX
PT    New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT    vasoconstriction
XX
PS    Disclosure; Page 37; 120pp; English.
XX
CC    The specification describes antisense oligonucleotides (AAX52869-X55271)
CC    directed against at least 2 mRNAs selected from target genes, coding and
CC    non-coding regions of RNAs corresponding to target genes, gene
CC    initiation codons, genomic flanking regions, intron-exon borders, the
CC    5'-end, the 3'-end and the juxta-section between coding and non-coding
CC    regions and all segments of RNAs encoding proteins associated with one
CC    or more diseases, conditions or mixtures. The antisense oligonucleotides
CC    may be derived from sequences AAX55272-74. These multiple target
CC    oligonucleotides (specifically AAX55180-271) can be used for the
CC    antisense treatment of diseases and conditions. Typical diseases and
CC    conditions are those associated with impaired respiration and
CC    inflammation, including lung diseases, pulmonary vasoconstriction,
CC    inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
CC    respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC    pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC    obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC    lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC    pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC    hepatic metastases, as well as all types of cancers which may metastasize
CC    or have metastasized to the lungs, including breast and prostate cancer.
XX
SQ    Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match             3.9%;  Score 58.2;  DB 20;  Length 114955;
Best Local Similarity  29.7%;  Pred. No. 0.0032;
Matches  204;  Conservative   76;  Mismatches  406;  Indels     0;  Gaps     0;

Qy     471 gaggctctccagcagcaggcgctggactggaccaaaatggcaactccaagcgtgctttgg 530
           |||||||||||||       :|   |||   |                     ::|||
Db  105181 GCGGGCGCGCCCCVNNHNNNNSCGGCCCGCGCGCCCVNNHNNNSCGGCCCCGGCC 105122

Qy     531 gggacccctgccactccacaggaagtctctcgcgccacctccgtcctcctgtcggc 590
           ||:||||   :|   ||||| ||                 ::   :||||||||||
Db  105121 GGCGGCGCCCCVGCNNHNNNSCGGCCCCGCGCCCVGGCCNNHNNNSCGGCC 105062

Qy     591 tcagcatggccgcgccgtcacaaactctgtagccctagccaggccaacaacagtg 650
           :|||||||||||||||      |    :::|   | |||||||| |
Db  105061 CGGCCGGCGGCGCGCGCCCVGCCNNHNNNSCGGCCCCVGGCCCCVVGGCCVNNH 105002

Qy     651 aaggcagcagaccatgactacctgccctggtgcggctggaggcaccaggctccttcc 710
           |::||:|||   ||||| ||||:|  |||||| |  |   |
Db  105001 NNNSCGGCCCGCGGCCGGCGCCCVGCCVGGCCVGNNHNNNSCGGCCGGCGGCCGCGC 104942

Qy     711 gcctggacggcgccttctcgccgcgtgtgccatctcgcaggagcgcctctgccgtgcct 770
           :|||||||| |  ||||:|   | :| |||||  |||||||||||| :|    |  |||
Db  104941 CCVGGCCVGCNNHNNNSCGGCCGGCCCVGGCCGGCGCGGCGCGCCCVGGCCVGCGCNNHNNNSGGCCVNNH 104882

Qy     771 cgcccttgcgtgaccggcaagcctcagcccacctggcctcattgcctggcgac 830
           :||   :|| | :||||||| :::|||| | ||||||||||||||||||||||
Db  104881 GCGGNNHNNNSVGGCCVGCGGNNHNNNSCVGGCCVGCGGNNHNNNSCCVGGCCVGCGGNN 104822

Qy     831 ttctcggccctggttgccaggggtgacagtgacacaggtggcacgccgtccccca 890
           ||:|:| |||||||| |||  |||||||||||||||| |||:|||:|:||||||||
Db  104821 HNNNSCCCVGGCCVGCGGNNHNNNSGCCCVGGCCVGCGGNNHNNNSGCCCCVGCCG 104762

Qy     891 gctcggggtctgagggccaggtgacagtggacagtggcggggctcagacaccgaggcct 950
           |:|:|||  ||||||||||||:|||:|  |||||||||| ||:||||:||||||||||
Db  104761 GNNHNNNSGCCCCVGGCCVGCGGNNHNNNSCGGCCCCVGGCCVGCGGNNHNNNSGCCCG 104702

Qy     951 ccgatggggcggaagggctgggcggcaccgacctgggcggcactgccg 1010
           ||||||||   ||||||| :|||:|  || |:| |||| ||||:||||
Db  104701 CCCVGCCVGCGGNNHNNNSGGGGCGCCCCVGGCCVGCGGNNHNNNSCCGGCGCCCCVGGCC 104642
```

CC　　pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC　　hepatic metastases, as well as all types of cancers which may metastasize
CC　　or have metastasized to the lungs, including breast and prostate cancer.
XX
SQ　　Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match　　　　　　3.7%;　　Score 55.4; DB 20;　Length 114955;
Best Local Similarity　33.0%;　Pred. No. 0.013;
Matches 217;　Conservative 37; Mismatches 403;　Indels　　0;　Gaps　　0;

Qy　　701 ggctccttccgcctggacgcgcccttcgcccgctgtgcatctcgcaggagcgcctc 760
　　　　　　|||　||||||||||||||　：　|||||||||||||||||||||||||||||||||
Db 108434 ggcggcgcctgctcgsnnndnnggcggggcggcggccgcctggctcsnnndnnggcggg 108493

Qy　　761 tgccgtgcctcgccccttgccgtgcacgcgccagcctcagcccaacctcggcctcattg 820
　　　　　　||：|||||||||||||||||||||　：　|||||||||||||||||||||||||||||
Db 108494 ggcggcggcgcctggctcgsnnndnnggcggggcggcggccgcctcgcsnnndnnggcggg 108553

Qy　　821 ccctggcacttctggccctggttgttggccaggtgacagtgccacggcctctgcagc 880
　　　　　　|　|||||　：||||||||||||||||||||||||||||||||||||||||||||
Db 108554 gcggcggccgcctggsnnndnnggcggggcggcggccgcctgsnnndnnggcggggcggc 108613

Qy　　881 ccgtccccagctcgggctctgagggcggccaggtggacagtggcagggctcagac 940
　　　　　　||：　||||||||||||||||||||||||||||||||||||||||||||||||
Db 108614 ggcgcctsnnndnnggcggggcggcggccgcsnnndnnggcggggcggcggccgcsnnn 108673

Qy　　941 accgaggcctccgatggggcggaaggcctggggcggcaccgacctgcggggccggacctgg 1000
　　　　　　|||　|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108674 dnnggcggggcggcggccgsnnndnnggcggggcggcggccsnnndnnggcggggcggc 108733

Qy　　1001 gccactgcctagcactgcctgctgctggagcaccgatgcgccgsnnndnnggcgacgagtg 1060
　　　　　　|：：　|||||||||||　　||　|||　：　||||||||||||||||||||||||||
Db 108734 ggsnnndnnggcggggcggcgsnnndnnggcggggcggcggccsnnndnnggcggggcggs 108793

Qy　　1061 gaactgacagcggccaagctgattgctgtctgcgggcccagcacttgctgacggcctt 1120
　　　　　　|||||||||||||| ||||||||||| ||||||||||||||：：||||||||||||
Db 108794 nnndnnggcggggcggcggccgcctgsnnndnnggcggggcggcggcggccsnnndnnggcggg 108853

Qy　　1121 gacctggccgcctcaaggccgagccgaggggctcttcctgctactgcgccactgggac 1180
　　　　　　||||||||||||||||　　||||||||| |||||||||||||||||||||||||||
Db 108854 cgcctggtctcgcctbgggcccsnnndnnggggcggcggcggccctbgggc 108913

Qy　　1181 caaaacctgcagctacacctgctgtgctacacagcccaggcgaacgtgtgaaggctgcccct 1240
　　　　　　|||　||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108914 ccsnnndnnggcggggcggccctbgggcccsnnndnnggcggggcggcggc 108973

Qy　　1241 gctgcttgggctggctgcgccccaccccaacactcaagtcactgcgccagggctggcctc 1300
　　　　　　|||||　　：|| ||| ||||| |||||||||||||||||||||||||||||||||
Db 108974 ctggctcgcctbgggcccsnnndnnggcggcggccctbgggcccsnnn 109033

Qy　　1301 ttggtgctggaaagtgtaggctggtgccagcctgtcccccactgcttccttactccc 1357
　　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109034 dnnggcggcggcggccctgctcgcctbgggcccsnnndnnggcggcggctggctccgc 109090

RESULT 9
AAA02477/c
ID　AAA02477 standard; cDNA; 1127 BP.
XX
AC　AAA02477;
XX
DT　19-MAY-2000　(first entry)
XX
DE　Human colon cancer cell line polynucleotide sequence SEQ ID NO:2468.
XX
KW　Human; colon cancer; tumour; diagnosis; gene expression product;
KW　probe; detection; cancerous state; metastasis; identification;
KW　breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW　oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS　Homo sapiens.
XX

Qy　　1011 tagcactcgcctggctggacgcaccgatgcgccgctccgacgagtgggaactgacag 1070
　　　　　　：|||　|||| |||：　　|：　：|||||||||||||||||||||||||||||||
Db 104641 CVGCGGNNHNNNSGCGGCGCCCCVGGCCVGCGGNNHNNNSGCGGCGCCGCCCVGGCCVG 104582

Qy　　1071 cggccaaggctgattgctgtgctcgggggccagcacttgcctgacggcctgactggccg 1130
　　　　　　|||　||||||||||||　　　||||||||||||||||||||||||||||||||||
Db 104581 CGGNNHNNNSCGGCGGCGCCCCVGCCGGCGGCGCCCVGCCGGNNHNNNSCCGCGGCGGCCGCGCCCCVGGCCV 104522

Qy　　1131 ccctcaaggccgagcccagggctc 1156
　　　　　　：　||||||||||||||||||||
Db 104521 GCGGNNHNNNSGCCCGGCGCGCCCC 104496

RESULT 8
AAX53491
AAX53491;
XX
DT　05-JUL-1999　(first entry)
XX
DE　Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW　Antisense oligonucleotide; multiple target; antisense treatment;
KW　impaired respiration; inflammation; lung disease;
KW　pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW　acute asthma; allergy; asthma; impeded respiration;
KW　respiratory distress syndrome; pain; cystic fibrosis;
KW　pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW　chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW　colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW　hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW　prostate cancer; ss.
OS　Synthetic.
XX
PN　WO9913886-A1.
XX
PD　25-MAR-1999.
XX
PF　17-SEP-1998;　98WO-US19419.
XX
PR　09-JUN-1998;　98US-0003972.
PR　17-SEP-1997;　97US-0059160.
XX
　　(UYEC-) UNIV EAST CAROLINA.
XX
　　Nyce JW;
XX
WPI;　1999-229400/19.
XX
PT　New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT　vasoconstriction
XX
PS　Disclosure; Page 37; 120pp; English.
XX
CC　The specification describes antisense oligonucleotides (AAX52869-X55271)
CC　directed against at least 2 mRNAs selected from target genes, coding and
CC　non-coding regions of RNAs corresponding to target genes, gene
CC　initiation codons, genomic flanking regions, intron-exon borders, the
CC　5'-end, the 3'-end and the juxta-section between coding and non-coding
CC　regions and all segments of RNAs encoding proteins associated with one
CC　or more diseases, conditions or mixtures. The antisense oligonucleotides
CC　may be derived from sequences AAX55272-74. These multiple target
CC　oligonucleotides (specifically AAX55180-271) can be used for the
CC　antisense treatment of diseases and conditions. Typical diseases and
CC　conditions are those associated with impaired respiration and
CC　inflammation, including lung diseases, pulmonary vasoconstriction,
CC　inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
CC　respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC　pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC　obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC　lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,

```
Db   587 CCNCNGCCCCNNCGNGACNNGGNGNGGCCCNNCCCCCGCNCCNNCCGGGGNC 528
                        |  |        |||        |     |  |     |
Qy  1105 cttgctgacggcctgacctggccgccgccctcaaggcgcagccgagggcttcctgct 1164
                  |||||  ||        |||||| ||
Db   527 NACGGCNCNGAGNCGGCNCNGNNNNNCCGNGGCCGCC--CCGNNCNCNNGGNCNCNCNC 470
            |     |        |  |        |   |        |  |
Qy  1165 actgcgccactggaccaaaacctgctgcagctacacctgctgtctacagccagcgaacgt 1224
                        |        |         |        |       |    ||
Db   469 NCNGCCCCGCGNNCNGNCNNCNNNNNACNNCNNGGNNNCCCCNNNCGNGNNCCGC 410
                  |    |        |  |      |   |   ||
Qy  1225 gtgaaggctgcccccctgctgcttgggctggccgNNNNNCCNNCCCCC 1284
               ||||   |      |||   ||      |     |
Db   409 GNNNNNCCNNCCCCCGNNNNNGCNNNNNGCCCCCCGGNNCCCCCCNNNCC 350
                     |        |    |  |        |||   ||
Qy  1285 gccaggggctggcctcttggtgctgggaaagtgtaggctcgtggtcgcagcctgtccccact 1344
               |   |        |        |  |   || ||| |||
Db   349 CGGCNNCNNNNNCCNNNCCNNNNCCCGNGNNNNNCGGNGNNNNNCGNNCCCCGNCNCNCNCC 290
                  |        ||| |||  |
Qy  1345 gcttcttactcctcccctcggagccctcttgccccacaaaaagtgctgctgtgtctc 1404
                  |||        |    |||        |     |    ||
Db   289 CCCNNNNNCGNNNNNNCNNCNNCNGNNNCNCNCNGCCCNNNNNNNCNCCCCCCGGNGNCC 230
                        ||        ||||   |
Qy  1405 tccctctctccaccccactcac 1428
                  |||  |    |||
Db   229 GGGCNCCGCGNCNCNGNCACNCCC 206

RESULT 10
AAK65257/c
ID   AAK65257 standard; DNA; 33718 BP.
XX
AC   AAK65257;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20069.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2001; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
```

```
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999; 99WO-US10602.
XX
PR   14-MAY-1998; 98US-0085426.
PR   15-MAY-1998; 98US-0085537.
PR   15-MAY-1998; 98US-0085696.
PR   21-OCT-1998; 98US-0105234.
PR   27-OCT-1998; 98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI   Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI   Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DB   WPI; 2000-125369/11.
```

Polynucleotide library used to determine cancerous states of mammalian
cells -

```
PS   Claim 1; Page 990; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived.
CC   The polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of
CC   pre-metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer.
XX
SQ   Sequence 1127 BP; 3 A; 214 C; 505 G; 49 T; 356 other;
```

```
Query Match             3.5%;  Score 53;  DB 21;  Length 1127;
Best Local Similarity   31.0%;  Pred. No. 0.017;
Matches  212; Conservative   0; Mismatches 470; Indels   2; Gaps   1;

Qy   745 ctcgcaggagcgcctctgccctgccgtgccgctgccctttgccgtgacccgacctcagccc 804
                  | -|  ||-||  ||    |  |||        ||  ||| ||
Db   887 CCNGNGGGNGCCCCCNCNGCNGCNCCCGNAGNGGCCNCCCACCGGCCNCCCCCNCNNGC 828
                  | |||||||  |||||   |
Qy   805 cacctcggcctcattgcctggcacttctgggccctggtgttggccagggtgacagtgc 864
                  |  ||  ||||||
Db   827 GCCCCCCNGNCNCCNCCNGTGCNNGNNCCCGNGNNCCCGCNCCGCCGNNGCNGTCCCCCC 768
                  |        |   |        |||  ||
Qy   865 cacggcctcctgcagccgcgtccccagctcgggcctcgggcctcgagggggcaggtgacag 924
                  |- |        |     ||
Db   767 NCCNGNGGCGCCCCCNNNGNCCCCGCNCGNACCCNCNNCGCGCCNCNNNGNCCCGCCCGCCCC 708
                  |   |        |||   |
Qy   925 tgggcggggctcagacaccgaggggcctcgatgggggcggaaagggctgggcgtggccgccacggacct 984
                  || ||        |||  |
Db   707 NCNNCNCGCGGNCNNCNCGCCNCNCCACGCGCCCNNNCGCGCCCCCNNNGNCCCNNCNGNCCC 648
                  ||| |        ||   |
Qy   985 gcgggccggacctgggcactgcctagcactcgctgctggagcaccgatgcggcgc 1044
                  || |  ||||||
Db   647 GGGGGGCCGCCNCCGCCCNCNCCCCNNCNCCCCCGNCNCGCGNNCNCCGGCNCCNCCGGCNCNNNNNCCCCCC 588
                  |   |        |||  |
Qy  1045 tgccttgcgagtgggaactgacagcggccaaggctgattgctggctggggcccagca 1104
```

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 20069; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 33718 BP; 9148 A; 6543 C; 7237 G; 10790 T; 0 other;

```
Query Match       3.3%;  Score 50;  DB 22;  Length 33718;
Best Local Similarity  50.0%;  Pred. No. 0.17;
Matches 181;  Conservative   0;  Mismatches 175;  Indels   6;  Gaps   2;

Qy   548 acggaagtcctcgcgccacctcccgtcctcccgtcggctcagcatgggcgccgt 607
            || || |  || -|- |   || ||||| |- |- ||| |   || | |||
Db   475 ACCCAATCGCGCGCGCGCCGCTCTTCCTCCTCCGCCCTCGTCGTCGTCTTCGT 416

Qy   608 cacaaactctgtagccctgacccgggccaggccaacaacagtgaaggcagcgaccatgac 667
            -|-| ||| || -| -|-| |  | -|-   || -| |||| | |||
Db   415 CGTACAAGTTGTCGGCCAGCGGTGGCTCCCGAGATGGGCCCTGCAGGCCCGGCCCCATTGT 356

Qy   668 tacctgccttggtgcggctgcaggaggcaccaggctccttccgcctggacgcgcccttc 727
            ||||||  -|-||-|-||  || |-  ||- || || -|-| || -| -|
Db   355 CTCCTTCCCCAGCCGCCGCAGTCG-----CCTGGGCCTCTTGCTCCCCGCCTGCCGCCGC 301

Qy   728 tgcgccgctgtgcgcatctgcaggagcgcctctgccgtgcctgcccttttgccgtgcac 787
            |||| -| || -| |||||| |-| |  |-||- -| - -|  | ||
Db   300 CGCCTCTGCCTCCGCCTCCGCCACAGCGCCGCCGCCACCCGGCAGCCCCTGGC 241

Qy   788 cgcgccagctcagccccacctcggcctcattgccctgggccctggtgtt 847
            ||| -|-|| |||-|- ||-|| ||-| |- -| ||| || ||| |-|-
Db   240 CGCCGCGGCCACCTCACGCTCTGGGGC-CGCCCCACCGGGCTCGCGGGCTCCGCTCGA 182

Qy   848 ggcagggtgacagtgccacggcctcctgcagcccgtcccccagctcgggctctgaggg 907
            ||-|-|||-|-|||  ||||-|-|-|-| |-|||| ||| |-||-|- ||| |-| -|
Db   181 GGCCGGGACCATCTCTCCCGGCCTCTTGCGGAGCGGCTCCCCGGGGGGGACGACGCGG 122

Qy   908 cc 909
            ||
Db   121 CC 120
```

Search completed: September 19, 2002, 04:08:14
Job time: 10924 sec

```
          GenCore version 4.5
     Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     September 19, 2002, 00:46:05 ; Search time 1848.71 Seconds
                                        (without alignments)
                                        11031.421 Million cell updates/sec

Title:           US-09-786-136-4
Perfect score:   1511
Sequence:        1 tccagagtgctgaatttctg........atccatgcaataagtgca 1511

ng table:        IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:        27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries
```

Database :   EST:

```
 1: em_estba:*
 2: em_esthum:*
 3: em_estin:*
 4: em_estmu:*
 5: em_estov:*
 6: em_estpl:*
 7: em_estro:*
 8: em_htc:*
 9: gb_est2:*
10: gb_htc:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB	ID	Description
1	596.6	39.5	841	10	BE798146	BE798146 601582424
2	560	37.1	561	10	BI793058	BI793058 1e47c12.y
3	523.8	34.7	527	10	BM353750	BM353750 ig55f05.y
c 4	472.8	31.3	495	10	BM127861	BM127861 if10c05.x
c 5	360.8	23.9	600	10	BI990349	BI990349 4064-13 M
c 6	319.2	21.1	333	10	BM353476	BM353476 ig55f05.s
c 7	317.2	21.0	643	12	BH051641	BH051641 RPCI-24-2
8	311	20.6	494	9	AW246364	AW246364 2821950.5
c 9	298.2	19.7	322	9	AW245580	AW245580 2821950.3
c 10	294	19.5	321	10	BI792808	BI792808 1e47c12.x
11	293.6	19.4	649	10	BB307349	BB307349 BB307349
12	207.8	13.8	526	10	BE808293	BE808293 213460 MA
13	195.8	13.0	286	10	T75425	T75425 yc90g11.r1
	170.8	11.3	470	10	BF412161	BF412161 UI-R-BT1-
	157	10.4	535	10	BE266573	BE266573 60119567
	154	10.2	423	10	BF550819	BF550819 UI-R-C1-j
	.9	9.9	425	9	BB711674	BB711674 BB711674

18	140	9.3	392	9	BB802514	BB802514 BB802514
19	139.8	9.3	348	10	BE862588	BE862588 UI-M-BH0-
20	130.6	8.6	371	9	BB804715	BB804715 BB804715
21	130.4	8.6	363	9	BB798951	BB798951 BB798951
22	130.4	8.6	410	9	AI044127	AI044127 UI-R-C1-}
c 23	128.2	8.5	411	9	AI547396	AI547396 UI-R-C3-t
c 24	124.4	8.2	917	12	AL275379	AL275379 Tetraodon
c 25	120.6	8.0	202	9	CNS045H6	CNS045H6
c 26	111.8	7.4	387	9	BB594108	BB594108 BB594108
27	111	7.3	312	10	BB800911	BB800911 BB800911
c 28	101.4	6.7	406	9	BF544431	BF544431 UI-R-BT0-
29	99.8	6.6	130	10	AI797955	AI797955 we83d10.x
30	94.8	6.3	526	9	BF189594	BF189594 235223 MA
c 31	92.2	6.1	316	9	AI145213	AI145213 UI-R-BT0-
					AI851743	AI851743 UI-M-BH0-
32	85	5.6	3760	11	AK016425	AK016425 Mus muscu
33	81.2	5.4	279	9	BB193885	BB193885 BB193885
34	79.6	5.3	280	9	BB180183	BB180183 BB180183
35	78.2	5.2	311	9	BB317839	BB317839 BB317839
c 36	77	5.1	505	9	AA630176	AA630176 ab95h06.s
37	74.2	4.9	323	9	BB182430	BB182430 BB182430
38	72.4	4.8	925	12	AL053013	AL053013 Drosophil
c 39	70.8	4.7	1651	10	BG809816	BG809816 mgct001xk
c 40	68.2	4.5	289	9	BB270582	BB270582 BB270582
c 41	66.6	4.4	935	12	CNS006XK	CNS006XK
42	66.6	4.4	932	12	AL066051	AL066051 Drosophil
43	65.2	4.3	222	9	CNS0072Q	CNS0072Q
44	64.6	4.3	221	9	BB309157	BB309157 BB309157
45	64.6	4.3	223	4	BB193279	BB193279 BB193279
					BB716322	BB716322 Mus muscu

ALIGNMENTS

RESULT 1
BE798146
LOCUS   BE798146   841 bp   mRNA   linear EST 20-SEP-2000
DEFINITION 601582424F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3937013 5',
            mRNA sequence.
ACCESSION   BE798146
VERSION   BE798146.1   GI:10219344
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 841)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM781 row: p column: 06
            High quality sequence start: 38
            High quality sequence stop: 795.
FEATURES   Location/Qualifiers
     source   1..841
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3937013"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
```

REFERENCE
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
          Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
          Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
          Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
          ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
          , Jackson,Y. and Bowers,Y.
TITLE     Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Other_ESTs: 1e47c12.x1
          Endocrine Pancreas Consortium
          Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
          Harvard University, Howard Hughes Medical Institute
          Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
          MA 02138
          Tel: 617-495-1812
          Fax: 617-495-8557
          Email: dmelton@biohp.harvard.edu
          Library was constructed by Dr. Douglas Melton DNA sequencing by:
          Washington University Genome Sequencing Center For information on
          obtaining a clone please contact: Juliana Brown
          (brown@fas.harvard.edu)
          Seq primer: -40RP from Gibco
          High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1..561
                     /organism="Homo sapiens"
                     /db_xref="taxon:9605"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT       86 a      213 c      154 g      107 t      1 others
ORIGIN

Query Match          37.1%;     Score 560;     DB 10;     Length 561;
Best Local Similarity   99.8%;     Pred. No. 1.4e-99;
Matches  560;     Conservative   0;     Mismatches   1;     Indels   0;     Gaps   0;

Qy    305  ttcacttgccctgtagctgtggatgctactactaggaggtcctgcctggggccctgcag  364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGAGGTCCTGCCTGGGGCCCTGCAG   60

Qy    365  gtgtgcagctcagagccgcctgagcccccaggaaccccctcctgcctctcacagccatcta  424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GTGTGCAGCTCAGAGCCGCCTGAGCCCCCAGGAACCCCCTCCTGCCTCTCACAGCCATCTA  120

Qy    425  gatgcagctcctctgccactgttgtctactctaaaggacttcagagaggctctccagca  484
           |||||||||||||||||||||||||  |  ||||||||||||||||||||||||||||||
Db    121  GATGCAGCTCCTCTGCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCA  180

Qy    485  ggcgctgggactcggaccaaaatggcaactccaagcgtgctttgggggacctgccact  544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GGCGCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCTGCCACT  240

Qy    545  cccacggaagtcctcgccgccacctcccgtcctccctgtcggctcagcatggcccgc  604
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

BASE COUNT      172 a     233 c     242 g     194 t
ORIGIN

          insert size 1.8kb. Library constructed by Ling Hong in
          the laboratory of Gerald M. Rubin (University of
          California, Berkeley) using ZAP-cDNA synthesis kit
          (Stratagene) and Superscript II RT (Life Technologies)."

Query Match          39.5%;     Score 596.6;     DB 10;     Length 841;
Best Local Similarity   97.0%;     Pred. No. 1.1e-106;
Matches  641;     Conservative   0;     Mismatches  14;     Indels   6;     Gaps   3;

Qy      1  tccagagtgctgaattcctgggcagccagggggctcttgctgctatgggtgaagatt   60
           ||||||||||||||||||||||||||||||||  |||||||||| ||||||||||||
Db    165  TCCAGAGTGCTGAATTTCTGGGCAGCCAGCCA-GGGGCTCTTGCTTGCTCTGCTATGGGTTGAAGATT  223

Qy     61  cagcctgatcacttcctgtcaaagttgggagtctgagctggttctggggcctagcaaggc  120
           ||||||||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db    224  CAGCCTGATCACTTCTTGTCAAAGTTGGGAGTCTGAGCTGGTTCTGGGGCCTAGCAAGGC  283

Qy    121  aattcactctgttaaagtcgataggaggagaagctgggagaacactgcaatttttcta  180
           |||| |||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    284  AATCTACTCTGTTAAAGTCGATAGAGAGGAGAAGCTGGGAGAACACTGCAATTTTTCTA  343

Qy    181  tgagccctgtagagggtgggtgggtggggggctccacagactgctcctggttccacagg  240
           |||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    344  TGAGCCCTCTGTAGAGGGTGGGTGGGTGGGGGGCTCCACAGACTGCTCCTTGTTCCACAGG  403

Qy    241  ccatgcccggaggtgcttggcttcgagccctcagacaagtaaggtcagctctgcccctc  300
           ||||||| |||||||||  ||||||||||| ||||||||||||||||||||||| ||||||
Db    404  CCATGCCCCGGAGGTGCTGGCTTCGAGCCCTTCAGACAAGTAAGGTCAGCTCTGCCCCCTC  463

Qy    301  ctgcttcacttgccctgtagtgtggatgtgctactactaggaggtcctgcctgggggccct  360
           |||| |||||||||||||||| |||||||| |||||||||||||||||||||||||  ||||
Db    464  CTGTCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGAGGTCCTGCCTGGGGCCCT  523

Qy    361  gcaggtgcagctcagagccgcctgagcccccaggaaccccctcctgcctctcacagca  420
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db    524  GCAGGTGTGCAGCTCAGAGCCGCCTGAGCCCCCAGGAACCCCCTCCTGCCTCTCACAGCCA  583

Qy    421  tctagatgcagctcctctgccactgttgtctactctaaaggacttcagagaggctctcc  480
           |||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Db    584  TCTAGATGCAGCTCCTCTGTCTCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCC  643

Qy    481  agcaggcggctgggactcggaccaaaatggcaactccaagcgtgctttgggggaccctgc  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    644  AGCAGGCGGCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCTGC  703

Qy    541  cactcccaccggaagtcctcgccgccacctcccgtcctccctgtcggctcagcatggg  600
           ||||||||||| ||  ||||||||||||||| ||  |||||||||||||||||||||||
Db    704  CACTCCCACCGAGGAGGTCCTCGCCGCCACCTCGCC--ACCTCCCGTCCTCCCTGTCGGCGCAGCATGGG  761

Qy    601  ccgcgtcacaaacctctgtagcctgacccggccaggccaacaacagtgaaggcagcga  660
           ||||| ||  |||||||||  |  ||||| | ||||||  |||| |||||||||||||||
Db    762  CCGCCGTCACAAAACTCTGTTAGCCTGACCGGG---CAGGCCCACAACAGTGAAGGCAGCGA  818

Qy    661  c  661
           -  -
Db    819  C  819

RESULT   2
BI793058
LOCUS        BI793058              561 bp    mRNA      linear   EST   01-OCT-2001
DEFINITION   1e47c12.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
             cDNA 5', mRNA sequence.
ACCESSION    BI793058
VERSION      BI793058.1  GI:15820783
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
Db    241 CCCACGGAAGGTCCTCGCCGCCCACCTCCCCGTCTCCCCGTCCCGTCGGCTCAGCATGGGCGC  300

Qy    605 cgtcacaaactcctgtagcctagccccggccaggccaacaacagtgaaggcagcagcgaccat  664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGTCACAAACTCTGTAGCCTCCTGACCCCGGCCAGCCCAACAACAGTGAAGGCAGCGACCAT  360

Qy    665 gactacctgccctggtgcggctgcaggaggcaccaggctccttccgcgcctggacgcgccc  724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GACTACCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGGGCTCCTTCCGGCGCCTGGACGCGCCC  420

Qy    725 ttctggcgccgtcgtggcatctgcaggagcgctctgccgtggctctgcccctttgcctg  784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTCTGGCGCCGTCGTGCCGCATCTGCAGGAGCGCCTCTGCCGTGCCTGCCCCTTTGCCGTG  480

Qy    785 caccgcgccaggcctcagccccacctcggcctcattgcctggcacctctgggccctggt  844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CACCGCGCCAGGCCTCAGCCCCACCTCGGCCTCATTGCCCTGNGCCACTTCTGGCCCTGGT  540

Qy    845 gtggccaggggtgacagtgcc  865
          |||||||||||||||||||||
Db    541 GTTGGCCAGGGGTGACAGTGCC  561
```

RESULT 3
BM353750

LOCUS　　　　BM353750　　　　　　　527 bp　　mRNA　　　linear　　EST 07-JAN-2002
DEFINITION　1g55f05.y1 HR85 islet Homo sapiens cDNA 5', mRNA sequence.
ACCESSION　BM353750
VERSION　　BM353750.1　　GI:18086108
KEYWORDS　EST.
SOURCE　　human.
ORGANISM　Homo sapiens
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1 (bases 1 to 527)
AUTHORS　　Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
　　　　　Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
　　　　　Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
　　　　　Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
　　　　　,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
　　　　　, Jackson,Y. and Bowers,Y.
TITLE　　　Endocrine Pancreas Consortium
JOURNAL　Unpublished (2000)
COMMENT　Other_ESTs: ig55f05.x1
　　　　　Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
　　　　　Endocrine Pancreas Consortium
　　　　　Harvard University, Howard Hughes Medical Institute
　　　　　Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
　　　　　MA 02138
　　　　　Tel: 617-495-1812
　　　　　Fax: 617-495-8557
　　　　　Email: dmelton@biohp.harvard.edu
　　　　　Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
　　　　　Washington University Genome Sequencing Center For information on
　　　　　obtaining a clone please contact: Dr. Hiroshi Inoue
　　　　　(hinoue@im.wustl.edu)
　　　　　Seq primer: ~40RP from Gibco
　　　　　High quality sequence stop: 444.
FEATURES　　　　　Location/Qualifiers
　　source　　　　1..527
　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　/clone_lib="HR85 islet"
　　　　　　　　　/tissue_type="Purified pancreatic islet"
　　　　　　　　　/lab_host="DH10B"
　　　　　　　　　/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
　　　　　　　　　NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
　　　　　　　　　Size-selected on agarose gel. Average insert size ~1kb. 5'
　　　　　　　　　XhoI site was destroyed after directional cloning.
　　　　　　　　　Amplified once. Contact information: Hiroshi Inoue, MD,
　　　　　　　　　Metabolism Div. (Alan Permutt Lab), Washington University
　　　　　　　　　School of Medicine, Box 8127, 660 South Euclid Ave., St.
　　　　　　　　　Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu; Tel:

BASE COUNT　　　74 a　　193 c　　178 g　　82 t
ORIGIN　　　　　314-362-1916, Fax: 314-747-2692."

Query Match　　　　34.7%;　　Score 523.8; DB 10;　Length 527;
Best Local Similarity　99.6%;　　Pred No. 1.6e-92;
Matches　525;　Conservative　0;　Mismatches　2;　Indels　0;　Gaps　0;

```
Qy    588 ggctcagcatggcgccgtcacaaatctgtagccctgaccccggccaggccaacaaca  647
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCTCAGCATGGGCGCCGTCACAAACTCTGTAGCCTCCTGACCCCGGCCAGGCCAACAACA  60

Qy    648 gtgaaggcagcgaccatgactacctgcccttggtggctggtcaggaggagcaccaggctcct  707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTGAAGGCAGCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGGCTCCT  120

Qy    708 tccgcctggacgcgcccttctggcgccgtcgtgcgcatctgcaggagcgctctctgccgtg  767
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCCGCCTGGACGCGCCCTTCTGGCGCCGTCGTGCGCATCTGCAGGAGCGCCTCTGCCGTG  180

Qy    768 cctcgcccttgccgtgcaccgcgccagcctcagcccaccctcggcctcattgcctggg  827
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTCGCCCTTGCCGTGCACCGCGCCAGCCTCAGCCCCACCTCGGCCTCATTGCCCTGGG  240

Qy    828 cacttctggccctggtgttggccaggtgacagtgcacagcgccgctcctgcagccgtccc  887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACTTCTGGCCCTGGTGTTGGCCAGGGTGACAGTGCACGGCGCCTCCTGCAGCCGTCCC  300

Qy    888 ccagtcggctctgagggccaggccaggtgacagtgggcgggctcagacaccgagg  947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCAGTCGGGCTCTGAGGGGCCAGGCCAGGTGACAGTGGGCGGGGCTCAGACACCGAGG  360

Qy    948 cctccgatgggcggaaggctgggcggcaccgacctgcgggcggacctggccactg  1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCTCCGAGGGGCGGAAGGCTGGGCGGCACCGACCTGCGGGCCGGGACCTGGCCCACTG  420

Qy   1008 ccgtagcactcgcctgctgagcaaccgatgcgccgctgccttcgacgagtgggaactga  1067
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCGTAGCACTCGCCTGGCTGGAGCACCGATGCGCCGCGTGCCTTCGACGAGTGGGAACTGA  480

Qy   1068 cagcggccaggctgattgctggctgcgggccagcacttgctgac  1114
          |||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGCGGCCAGGCTGATTGCTGGCTGCGGGCCCCAGCACTTGCCTGAC  527
```

RESULT 4
BM127861/c

LOCUS　　　　BM127861　　　　　　　495 bp　　mRNA　　linear　　EST 27-NOV-2001
DEFINITION　1f10c05.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
　　　　　cDNA 3', mRNA sequence.
ACCESSION　BM127861
VERSION　　BM127861.1　　GI:17122413
KEYWORDS　EST.
SOURCE　　human.
ORGANISM　Homo sapiens
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1 (bases 1 to 495)
AUTHORS　　Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
　　　　　Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
　　　　　Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
　　　　　Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
　　　　　,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
　　　　　, Jackson,Y. and Bowers,Y.
TITLE　　　Endocrine Pancreas Consortium
JOURNAL　Unpublished (2000)
COMMENT　Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
　　　　　Endocrine Pancreas Consortium
　　　　　Harvard University, Howard Hughes Medical Institute
　　　　　Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
　　　　　MA 02138
　　　　　Tel: 617-495-1812

```
       Fax: 617-495-8557
       Email: dmelton@biohp.harvard.edu
       Library was constructed by Dr. Douglas Melton DNA sequencing by:
       Washington University Genome Sequencing Center For information on
       obtaining a clone please contact: Juliana Brown
       (brown@fas.harvard.edu)
       High quality sequence stop: 446.

FEATURES             Location/Qualifiers
     source          1..495
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."

BASE COUNT    95 a    120 c    183 g    97 t
ORIGIN

Query Match     31.3%;   Score 472.8;  DB 10;  Length 495;
Best Local Similarity 99.6%;  Pred. No. 1.5e-82;
Matches 474;  Conservative 0;  Mismatches 0;  Indels 2;  Gaps 0;

Qy  1035 gatgcgccgtgcctcgacgagtgggaactgacagcggccaaggctgattgctggctgc 1094
              ||||||||| ||||| |||||| |||||||||||||||||||||||||||||||||||||
Db   495 GATGCGCCGCTGCCTTCGACGACGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGGCTGC 436

Qy  1095 gggcccagcactgcctgacggcttgacctggccgccgtccaaggccgagcccgagggc 1154
              ||||||||||| |||||||||| |||||||||||||| |||||||||||||||||||||
Db   435 GGGCCCAGCACTTGCCTGACGGCCTTGACCTGGCCGCCGCCCAAGGCCGAGCCCGAGGGC 376

Qy  1155 tcttcctgctactgcgccactgggaccaaaacctgcagctacacctgtgtgctacagcc 1214
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   375 TCTTCCTGCTACTGCGCCACTGGGACCAAAAACCTGCAGCTACACCTGTGTGCTACAGCC 316

Qy  1215 cagcgaacgtgtgaaggctgccccctgctgcttgggctgggcgcccccaccaacactca 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315 CAGCGAACGTGTGAAGGCTGCCCCCTGCTGCTTGGGCTGGGCGCCCCCACCAACACTCA 256

Qy  1275 agtcactgccgcgccaggggctggcctctggtgctgtgggaaagtgtaggctggtgccagcct 1334
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   255 AGTCACTGCCGCGCCAGGGGCTGGCCTCTTGGTGCTGTGGGAAAGTGTAGGCTGGTGCCAGCCT 196

Qy  1335 gtccccactgcttcttactcccctcccctggagcccctcttgcccccacaaaagtgcctgc 1394
              ||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db   195 GTCCCCACTGCTCTTACTCCCTCCCCCTGGAGCCCCTCTTGCCCCCACAAAAAGTGCCTGC 136

Qy  1395 ctgtgtctctcctctctcccacccacccctcacactcccctccatcctctgagctccct 1454
              |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db   135 CTGTGTCTCTCCTCTCTCCCTCCCCCACCCACTCACACTCCCCTCCATCCTTTGAGCTCCCT 76

Qy  1455 gcaacacagtggaagggtagagagccacagtcccaaatcctatgcaataaagtgc 1510
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    75 GCAACACAGTGGAAGGGTAGAGAGCCCAACACCTGCCACAGTCCCAAATCCTATGCAATAAAGTGC 20

RESULT 5
BI990349
```

```
LOCUS       BI990349     600 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  4064-13 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BI990349
VERSION     BI990349.1  GI:17961359
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
AUTHORS     Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
TITLE       Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
JOURNAL     Nucleic Acids Res. 29 (24), 4983-4993 (2001)
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"

BASE COUNT    111 a    179 c    163 g    130 t    17 others
ORIGIN

Query Match     23.9%;   Score 360.8;  DB 10;  Length 600;
Best Local Similarity 79.4%;  Pred. No. 1.2e-60;
Matches 452;  Conservative 0;  Mismatches 114;  Indels 3;  Gaps 3;

Qy   665 gactaacttgccctggtgcggc-tgcaggaggagccacaggctccttccgcctggacgcgcc 723
              |||||| || ||| ||| | ||| ||| ||||| |||||||||||||| ||||||| || ||
Db     1 GACTACCTGCCTCTTGGTACGACGTGCAAGAGGCACCACAGGCTCCTTNCGCCTGGATGAACC 60

Qy   724 ctttctgccgcgctgtgcgatctcgcagagcgggaggcctctg--ccgtgctccgccccttgccg 782
              || |||   || ||| || |  |   |||| |   ||||  |||  | |||||| ||||||| | ||
Db    61 CTTCTGTGTGCTGNTGTGTGCATTCCTCAAGAACCGCCTGTGTGTNAGAGCCTCGCCCCTTTGTCG 120

Qy   783 tgcaccggcgccagcctcagcccacccggcctcattgccctgggcacttctgggccctg 842
              | ||||||| ||| |||||||||  ||  | ||| |  ||| ||||| ||||| |||||
Db   121 CACACCGGTGCCANCCTCAGCCCCCACCTNAGCNTCATNTCCTNGGCATTTCTGAGCCCTG 180

Qy   843 gtgttggcaggtgacagtgccacgcggcctcctgcagcccgtccccagctcgggcttctg 902
                |||||| ||||||||||||||| | ||||||||||||  |||||||||| |||  ||||
Db   181 NTATTGGCCAGGTGACAGTGCCACANCCTCCTGCAGCCAGTGCCCCAGCTCAGGTCTG 240

Qy   903 aggggccaggccaggtggacag-tggccgggggctcagacaccgaggcctccgatggggcg 961
              ||| |||||||||||| ||||| |||  ||||||||||| |  ||||||||| | ||  ||
Db   241 AGGGTCCAGGCCAGGTAGACAGCTGGGAGGGGCTCAGATACTGAGGCCTCATAAGGAATG 300

Qy   962 gaagggctgggcggcaccgacctgcgggggccgggccgactcggcactgccgtagcactcgcc 1021
              ||| ||| ||| ||||||| |||||| || ||  |  ||| | | |||||||| |   |||||
Db   301 GAAAGGCAGGACAGCCTGTGATCTTCTCNNGGGCGCACTTGGCCCACCACAGNTGNGGNCCTCGGCG 360

Qy  1022 tggctggagcaccgatgcgccgccgctgccttcgacgagtggggaactgacacgcggccaaggct 1081
              ||||||| ||||| |||| ||| | |||||||| | |||| || |||||||| ||  ||||| ||
Db   361 TGGCTGGAACACCCGCTGTGCAGCTGCCGCTGCCTTTGGCAGTGGCAGAACTGACAGCTTCCAAAGCT 420

Qy  1082 gattgctggctgcggggccccagcacttgcctgacggccttgacctggccgccccccaaggcc 1141
              |||||||||||| |  | || ||||| |||| ||||| ||||| |||| | |   |  |||
Db   421 GATTGCTGGCTGAGAGCCCAACACCTGCCAGACGGACTTGATCTGACTGCTTAAAGCT 480

Qy  1142 gcagcccgagggctccttcctgctactgcgccactgggaccaaaacctgcagctacacctg 1201
              || |||| |||||||| ||| ||| |  | ||||  | ||| |  | ||  | |||||| ||
Db   481 GCTGCCCGGGGGCTCTTCTCGTTCTGTTACTGCGACATCTGCGACGANTCTGCAGCTACACTTG 540
```

```
Qy   1202 ctgtgctacagcccagcgaacgtgtgaag 1230
          |||| ||| | |     ||  |  ||| |
Db    541 CTGTGTTACANCCCTTCGAATGTGTAANG 569

RESULT   6
BM353476/c
LOCUS       BM353476                333 bp    mRNA    linear   EST 07-JAN-2002
DEFINITION  1g55f05.x1 HR85 islet Homo sapiens cDNA 3', mRNA sequence.
ACCESSION   BM353476
VERSION     BM353476.1  GI:18085834
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 333)
AUTHORS     Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
TITLE       Endocrine Pancreas Consortium
JOURNAL     Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 264.
FEATURES             Location/Qualifiers
     source          1..333
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HR85 islet"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming. 5'
                     Size-selected on agarose gel. Average insert size -1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave. St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
BASE COUNT     69 a      70 c     130 g     64 t
ORIGIN

Query Match      21.1%;   Score 319.2;  DB 10;  Length 333;
Best Local Similarity  99.1%;   Pred. No. 1.4e-52;
Matches  321;  Conservative  0;  Mismatches  0;  Indels  3;  Gaps  0;

Qy   1187 ctgcagctacacctgctgtgctacagcccagcgaacgtgtgaaggctgcccccctgctgct 1246
          ||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||| ||
Db    333 CTGCAGCTACACCTGCTGTGCTACAGCCCAGCCGAACGGCTGTGAAGGCTGCCCCCTGCGCT 274

Qy   1247 tgggctggcggcgccaccaacacactcaagtcactgcgcgccagggctggcctccttggtg 1306
          |||||||||| ||||||||| |||||||||||||||||  |||| |||||||| ||| |||
Db    273 TGGGCTGGCGCCCCCACCCAACACACTCAAGTCACTGCCCCCCAGAACTGGCCTCTTGGTG 214

Qy   1307 ctgggaagtgtaggctggtgccagcctgtcccccactgcttcttactccctccctggag 1366
          ||| | ||||||||||||||||||||||| ||||||||||||||| |||||| ||| |||
Db    213 CTGGAAAGTGTAGGCTGGTGCCAGCCTGTCCCCCACTGCTTCTTACTCCTCCTAGAG 154

Qy   1367 ccctcttgcccccacaaaagtgctgctgtctctccctcctccacccaccactc 1426
          ||||||||||||||| |||||||||||||||||| || ||| |||||||||||
Db    153 CCCTCTTGCCCCCACCAAAAAGTGCTGCTGTCTCTCCCCTCTCCCCCACCCCACTC 94

Qy   1427 acactccctccactccttgagctccctgcaacacagtggaaggtagagagccacagtc 1486
          |||||||| ||||| |||||||||||||||||||||||||||| |||||||||||||||
Db     93 ACACTCCCCCTCCATCCTCTGAGCTCCCTGCAACACAGTGGAAGGGTGAGAGAGCCACAGTC 34

Qy   1487 cccaaatcctatgcaataaagtgc 1510
          ||||||||||||||||||||||||
Db     33 CCCAAATCCTATGCAATAAAGTGC 10

RESULT   7
BH051641
LOCUS       BH051641                643 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-260N14.TJ RPCI-24 Mus musculus genomic clone RPCI-24-260N14
            , DNA sequence.
ACCESSION   BH051641
VERSION     BH051641.1  GI:14843393
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
TITLE       Mouse BAC End Sequences from Library RPCI-24
JOURNAL     Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-260N14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 260  Row: N  column: 14
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..643
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-260N14"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1:
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI partially digested male C57BL/6J
                     DNA."
BASE COUNT    128 a     204 c     178 g     133 t
ORIGIN

Query Match      21.0%;   Score 317.2;  DB 12;  Length 643;
Best Local Similarity  80.3%;   Pred. No. 4e-52;
Matches  384;  Conservative  0;  Mismatches  93;  Indels  1;  Gaps  1;

Qy    661 ccatgactacctgccctggtgcggctgcaggaggcaccaggctccttccgcctggacgc 720
          |||| |||  ||| || || |||  ||| | ||| ||  || |||| | ||||||| |
```

```
            EcoRI; cDNA made by oligo-dT priming, Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G).  Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT     92 a    161 c    148 g     93 t
ORIGIN

Query Match         20.6%;   Score 311;   DB 9;   Length 494;
Best Local Similarity  100.0%;  Pred. No. 6e-51;
Matches  311;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   239 ggccatgccgaggtgctgcttcgagccctcagacagtaaggtcagctctgccccc 298
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 GGCCATGCCCGGAGGTGCTGGCTTCGAGCCCTCAGACAGTAAGGTCAGCTCTGCCCCC 243

Qy   299 tcctgcttcacttgcctgtagctgtggatgctactactagggagggtcctgcctggggcc 358
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 TCCTGCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGGAGGGTCCTGCCTGGGGCC 303

Qy   359 ctgcaggtgtgcagctcagagcgccgtgagccccccaggagaaccccctcgcctctcacagc 418
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 CTGCAGGTGTGCAGCTCAGAGCGCCGTGAGCCCCGGAGGACCCCGTGAGCCCCTCGCCTCTCACAGC 363

Qy   419 catctagatgcagctcctctgccactgttgtctactctaaaggacttcagagaggctct 478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 CATCTAGATGCAGCTCCTCTGCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCT 423

Qy   479 ccagacaggcagcctgggactcggaccaaaatggcaactccaagcgtgcttgggggacct 538
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 CCAGCAGGCGCCTGGGACTCGGACCTCGGACCAAAATGCAACTCCAAGCGTGCTTTGGGGACCCT 483

Qy   539 gccactcccac 549
         |||||||||||
Db   484 GCCACTCCCAC 494

RESULT 9
AW246580/c
LOCUS       AW246580                  322 bp   mRNA    linear   EST 07-JAN-2000
DEFINITION  2821950.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821950 3',
            mRNA sequence.
ACCESSION   AW246580
VERSION     AW246580.1  GI:6589573
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 322)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Other_ESTs: 2821950.5prime
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling
            Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
            Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
            project Clone distribution: MGC clone distribution information can
            be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
            Scores: PHRED from University of Washington Genome Center. Quality
            Trimming: cross_match from University of Washington Genome Center
            PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
            Drosophila Genome Project. University of Washington Genome Center:
            http://www.genome.washington.edu Polyadenylation: Based upon the
            presence of a XhoI site followed by a run of 14 or more T residues
            at the beginning of the sequence, this cDNA insert was
            polyadenylated.
```

```
Db   163 CCACCTCTTCTCTCCTCAGGTACGACTGCAAGAGGCACCAGGCTCCTTCCGCCTGGATGA 222
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721 gccttctgccgccgctgcgatctcgagggagcgcctcgccgtgcctgccccttgc 780
         ||||||| |||||||| |||||| ||||||||||||||||||||||||||||||||
Db   223 ACCCTTCGTGCTGCTGTGTGCATTCCTCAAGAACGCCTGTGTGCAGCCTGCCCCTTTGC 282

Qy   781 cgtgcaccgcgcagccagccaccctcggcctcattgcctgccgtggcgcactttggggccc 840
         |||| ||||||||||||||||||| |||||||||||| ||| ||||| |||||||||||
Db   283 CGCACACCCTGCCAGCCTGCCAGCCCCACCCCAGCATCACTCCTTGGGCATTTCTGAGCCCC 342

Qy   841 tgtgtgttggccaggtgacagtgcacacgccctctgagacgccgtccccccagctcggggctc 900
         |||| |||| |||| || |||| |||||||||||||| |||| |||||||| ||||||||||
Db   343 TGGTATTGGCCAAGGTGACAGTGCACAGTGCCCACAGCCCTCGAGCCAGCTCCCCCAGCGTCAGGTC 402

Qy   901 tgaggggccaggccagtggacagtgggcgggggctcagacaccgaggcctccgatggggc 960
         |||||||||||| |||||||||||||||||||||||||||| ||||||| ||||||||| ||
Db   403 TGAGGGTCCAGGCCAGGCCAGGTAGACACGTGGGAGGGGGTCAGATACTGAGGCCTCAGAAGGAAT 462

Qy   961 ggaagggctgggcgggcaccgacctgggcggcacctgcgtagcactgc 1020
         |||||||||||||||| ||||||| ||| ||||||| |||||||||||
Db   463 GGAAAGGCAGGACAGCTCYATCYTCGTGGGCGCACTTGGCCCACAGCTGTGGCCCTGGC 522

Qy   1021 ctggctgagcaccgatgcgccgctgccttcgacgagtggaactgacagcggccaaggc 1080
          ||||||||||||||||||||| |||||| |||||||||| |||||||||||||||||
Db   523 GTGGCTGGAACACCGCTGCGCAGCTGCCTTTGGCGAGTGGGAACTGACAGCTTCCCAAGC 582

Qy   1081 tgattgc-tggctgcgggccagacttgcctgacggcgccctcaa 1137
          |||||| |||||||||||||||||| ||||||||||||||||
Db   583 TGATTGCTTGGCCTGAGAGCCCTGAGACTGACTGACTCGTCTTAA 640

RESULT 8
AW246364
LOCUS       AW246364                  494 bp   mRNA    linear   EST 07-JAN-2000
DEFINITION  2821950.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821950 5',
            mRNA sequence.
ACCESSION   AW246364
VERSION     AW246364.1  GI:6589357
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Other_ESTs: 2821950.3prime
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling
            Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
            Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
            project Clone distribution: MGC clone distribution information can
            be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
            Scores: PHRED from University of Washington Genome Center. Vector
            Trimming: cross_match from University of Washington Genome Center
            PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
            Drosophila Genome Project. University of Washington Genome Center:
            http://www.genome.washington.edu
            High quality sequence stop: 320.
            Plate: LLCM8 row: C column: 7
FEATURES             Location/Qualifiers
     source          1..494
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2821950"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
```

```
          Plate: LLCM8  row: C  column: 7
          High quality sequence stop: 183.
FEATURES
     source    1..322
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:2821950"
               /clone_lib="NIH_MGC_7"
               /tissue_type="small cell carcinoma"
               /cell_line="MGC3"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally
               cloned into EcoRI/XhoI sites using the following 5'
               adaptor: GGCACGAG(G).  Size-selected >500bp for average
               insert size 1.8kb. Library constructed by Ling Hong in
               the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT     66 a      67 c     121 g      68 t
ORIGIN

Query Match              19.7%;   Score 298.2;  DB 9;    Length 322;
Best Local Similarity    99.0%;   Pred. No. 1.7e-48;
Matches  300;  Conservative   0;  Mismatches   3;  Indels   0;  Gaps   0;

Qy  1208  tacagcccagcgaacgtgtgaaggctgcccctgctgcttgggctggcgccccaccaac  1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322  TACAGCCCAGCGAACGTGTGAAGGCTGCCCCTGCTGCTTGGGCTGGCGCCCCACCCAAC   263

Qy  1268  acactcaagtcactgccgccagggctggcctcttggtgctgggaaagtgtaggctggtg  1327
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   262  ACACTCAAGTCACTGCCGCCAGGGCTGGCCTCTTGGTGCTGTGTAGGCTGGTT   203

Qy  1328  ccagcctgtccccactgcttcttactcctcctggagccctcttgccccaccaaaag  1387
          ||||||||||||||||||||||||||||| ||||||||||||||||||| ||||||||
Db   202  CCAGCCTGTCCCCACTGCTTCTTACTCCTCCCAGAGCCCTCTTGCCCCCACAAAAG   143

Qy  1388  tgcctgcctgctcctctccctcctcctccacccaccactccccTCCATCCTCTGA  1447
          ||||||||||||||||||||||||||||||||||||||||||| ||||||| |||
Db   142  TGCCTGCCTGCTCCTCTCCCTCCTCCTCCACCCACCACTCCCCATCCTCCTTTGA    83

Qy  1448  gctccctgcaacacagtggaaggtagagagccacagtcccaaatcctatgcaataaag  1507
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    82  GCTCCCTGCAACACAGTGGAAGGGTAGAGAGCCACAGTCCCAAATCCTATGCAATAAAG    23

Qy  1508  tgc  1510
          |||
Db    22  TGC    20

RESULT 10
BI792808/c
LOCUS       BI792808    321 bp    mRNA    linear    EST 01-OCT-2001
DEFINITION  1e47c12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3', mRNA sequence.
ACCESSION   BI792808
VERSION     BI792808.1  GI:15820533
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
```

```
COMMENT    Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
           Endocrine Pancreas Consortium
           Harvard University, Howard Hughes Medical Institute
           Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
           MA 02138
           Tel: 617-495-1812
           Fax: 617-495-8557
           Email: dmelton@biohp.harvard.edu
           Library was constructed by Dr. Douglas Melton DNA sequencing by:
           Washington University Genome Sequencing Center For information on
           obtaining a clone please contact: Juliana Brown
           (brown@fas.harvard.edu)
           Seq primer: -40UP from Gibco
           High quality sequence stop: 286.
FEATURES
     source    1..321
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
               /sex="Both"
               /tissue_type="Islets of Langerhans"
               /dev_stage="Adult"
               /lab_host="DH10B"
               /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
               Site_2: Sal 1; Starting library constructed using
               SuperScript Plasmid Library kit (Life Technologies). cDNA
               made by oligo-dT priming. Size-selected by column
               fractionation; average insert size 1.08 kb. Library was
               amplified once on solid support and plasmid DNA from
               library was prepared. The library DNA was normalized by
               method #4 from Bonaldo, Lennon, and Soares 1996 Genome
               Research 6:791-806: 0.5 microgram single-stranded library
               plasmid DNA was mixed with 5 micrograms PCR product
               representing library inserts and hybridized to an Ecot of
               20. Single-stranded (unhybridized) plasmids were isolated
               by hydroxyapatite chromatography and used to make this
               library."

BASE COUNT     67 a      69 c     119 g      65 t      1 others
ORIGIN

Query Match              19.5%;   Score 294;   DB 10;    Length 321;
Best Local Similarity    98.0%;   Pred. No. 1.1e-47;
Matches  297;  Conservative   0;  Mismatches   6;  Indels   0;  Gaps   0;

Qy  1208  tacagcccagcgaacgtgtgaaggctgcccctgctgcttgggctggcgccccaccaac  1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321  TACAGCCCAGCGAACGTGTGAAGGCTGCCCCTGCTGCTTGGGCTGGCGCCCCACCCAAC   262

Qy  1268  acactcaagtcactgccgccagggctggcctcttggtgctgggaaagtgtaggctggtg  1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261  ACACTCAAGTCACTGCCGCCAGGGCTGGCCTCTTGGTGCTGGGAAAGTGTAGGACTGGTG   202

Qy  1328  ccagcctgtccccactgcttcttactcctcctggagccctcttgccccaccaaaag  1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201  CCAGCCTGTCCCCACTGCTTCTTACTCCTCCTGGAGCCCTCTTGCCCCCACAAAAG   142

Qy  1388  tgcctgcctgctcctctccctcctcctccacccaccactccccTCCATCCTCTGA  1447
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141  TGCCTGCCTGCTCCTCTCCCTCCTCCTCCACCCACCACTCCCCATCCTCCTTTGA    82

Qy  1448  gctccctgcaacacagtggaaggtagagagccacagtcccaaatcctatgcaataaag  1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81  GCTCCCTGCAACACAGTGGAAGGGTAGAGAGCCACAGTCCCAAATCCTATGCAATAAAG    22

Qy  1508  tgc  1510
          |||
Db    21  TGC    19

Search completed: September 19, 2002, 03:21:01
Job time: 9296 sec
```

us-09-786-136-4.rst

```
                    GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.
```

OM nucleic - nucleic search, using sw model

Run on: September 19, 2002, 00:57:10 ; Search time 54.09 Seconds
                              (without alignments)
                              6861.753 Million cell updates/sec

Title: US-09-786-136-4
Perfect score: 1511
Sequence: 1 tccagagtgctgaattctg........atcctatgcaataaagtgca 1511

ng table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Issued_Patents_NA:*
            1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	59	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
2	49.8	3.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
3	49.8	3.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
4	46	3.0	15872	4	US-09-105-537-1	Sequence 1, Appli
c 5	45.6	3.0	4403765	4	US-09-103-840A-2	Sequence 32, Appli
6	44.2	2.9	11220	4	US-09-105-537-32	Sequence 32, Appli
7	44.2	2.9	36778	4	US-09-105-537-5	Sequence 5, Appli
8	44.2	2.9	38506	3	US-09-320-878-19	Sequence 19, Appl
9	44	2.9	2064	1	US-08-343-428-1	Sequence 1, Appli
10	43.6	2.9	3600	2	US-08-537-002A-5	Sequence 5, Appli
11	43.6	2.9	3600	3	US-08-863-010-5	Sequence 5, Appli
12	43.6	2.9	3600	4	US-09-024-429-5	Sequence 5, Appli
13	43.4	2.9	1860	3	US-08-331-644-3	Sequence 3, Appli
14	43.4	2.9	1860	5	PCT-US93-04102-3	Sequence 29, Appl
c 15	43.2	2.9	1776	1	US-08-722-001-29	Sequence 19, Appli
16	43	2.8	1288	5	US-08-440-856A-9	Sequence 9, Appl
17	43	2.8	1474	5	PCT-US94-00545-19	Sequence 1, Appli
18	43	2.8	3478	1	US-08-396-479B-1	Sequence 1, Appli
19	43	2.8	3478	1	US-08-818-823-1	Sequence 1, Appli
c 20	42.8	2.8	2823	1	US-08-398-008A-1	Sequence 1, Appli
c 21	42.8	2.8	2823	1	US-08-893-333-1	Sequence 1, Appli
c 22	42.8	2.8	4411529	4	US-09-103-840A-1	Sequence 4, Appli
23	42	2.8	2889	1	US-08-537-002A-4	Sequence 4, Appli
24	42	2.8	2889	3	US-08-863-010-4	Sequence 4, Appli
25	42	2.8	2889	4	US-09-024-429-4	Sequence 52, Appl
26	41.8	2.8	1162	2	US-08-726-306A-52	Sequence 8, Appli
	41.8	2.8	1480	2	US-08-839-008-8	
28	41.8	2.8	2923	1	US-08-377-292-6	Sequence 6, Appli
29	41.8	2.8	2923	4	US-07-989-847-7	Sequence 7, Appli
30	41.8	2.8	2923	2	US-08-469-411-7	Sequence 7, Appli
31	41.4	2.7	3382	2	US-08-682-847-1	Sequence 1, Appli
32	41.4	2.7	43280	4	US-08-804-227C-1	Sequence 1, Appli
c 33	41.2	2.7	780	4	US-09-323-555B-1	Sequence 2, Appli
c 34	41.2	2.7	53526	3	US-08-658-136-2	Sequence 7, Appli
35	41.2	2.7	53577	3	US-08-658-136-1	Sequence 1, Appli
36	41	2.7	44377	4	US-08-804-227C-7	Sequence 1, Appli
37	41	2.7	44377	2	US-08-804-198-1	Sequence 1, Appl
c 38	40.6	2.7	49136	4	US-09-422-869-1	Patent No. 5187076
39	40.2	2.7	1537	6	5187076-5	Sequence 9, Appli
40	40.2	2.7	2923	6	5187076	Sequence 1, Appli
c 41	40.2	2.6	1288	2	US-08-440-856A-9	Sequence 3, Appli
c 42	40	2.6	1402	1	US-08-447-965A-1	Sequence 3, Appli
c 43	40	2.6	1417	4	US-09-199-737-3	
c 44	39.8	2.6	1227	3	US-09-074-912-3	
c 45	39.8	2.6	1227	4	US-09-290-136-3	

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
;   FILING DATE:
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE:
;   APPLICATION NUMBER: EP 91 114 300.6
;   FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 7218 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   CLONE: pTvgpt-Fls
US-08-232-463-14
```

[handwritten annotation: "Not this relevant"]

Query Match   3.9%;   Score 59; DB 1; Length 7218;
Best Local Similarity 2.9%; Pred. No. 0.00011;
Matches 11;   Conservative 226; Mismatches 146; Indels   0; Gaps   0;

Qy   1072 ggccaaggctgattgctggctgcggggccagcacttgcctgacggcctggccgc 1131
          || || ||| || | :|| ::| :||: ::|:::
Db   1048 GGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1107

Qy   1132 cctcaaggccgcagcccgagggctcttcctgctactgcgccactggaccaaaacctgca 1191
          :::: ||| ||| || ||| YYYYYYYYYYYYYYYYYYYYYYYYYYY 1167
Db   1108 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

Qy   1192 gctacacctgctgctacagcccagcgaacgtgtgaaggctgccccctgcttgggc 1251
          :: :::::|: ::: : :||: YYYYYYYYYYYYYYYYYYYYYYYY 1227
Db   1168 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

Qy   1252 tgcgccccaccaacacactcaagtcactgccgccagggctggcctcttggtgctggg 1311
          ::::::|||: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1287
Db   1228 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

Qy   1312 aaagtgtaggctggtgcagcctgtccccactgcttcttactccctccctgagccctc 1371
          :::::::::|:: :::::|||: YYYYYYYYYYYYYYYYYYYYY 1347
Db   1288 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

Qy   1372 ttgcccccacaaaaagtgcctgctgtctctctccctctcctccaccccactcacact 1431
          : ::::|: ::::: YYYYYYYYYYYYYYYYYYYYY::::: 1407
Db   1348 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

Qy   1432 ccctccatcctctgagctccct 1454
          :: ::::: YYYYYYYYYYYYY 1430
Db   1408 YYYYYYYYYYYYYY

RESULT   2
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENTER, John C.
;   TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;   TITLE OF INVENTION: TUBERCULOSIS
;   FILE REFERENCE: 24366-20007.00
;   CURRENT APPLICATION NUMBER: US/09/103,840A
;   CURRENT FILING DATE: 1998-06-24
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match   3.3%;   Score 49.8; DB 4; Length 4403765;
Best Local Similarity 48.2%; Pred. No. 0.089;
Matches 172;   Conservative   0; Mismatches 182; Indels   3; Gaps   1;

Qy   695 gcaccaggctcctccgcctgacgcgccttcgcgcctgtgcgcatctcgcaggag 754
          || | || | || |:|| |||||| | ||||:| | ||| | |||||
Db 337106 gcgcgcggcgccgaacagcgccgtgcccccggcgcccgcgtcggcgcgagcgtgcc 337165

Qy   755 cgcctctgccgtgccgtcgcctttgccgtcaccgccgccagcctcagcccacctcggcc 814
          |||| | | :| ||||: | | :|||||| | | ||| ||| ||| || |
Db 337166 cgccccgccggtgccgcggcgcggcgcgaagagcaatcgttcccccggtcccgccattcgc 337225

Qy   815 tcattgccctgggcacttctgggcctgtgttggc---caggtgacagtgccacggcc 871
          | | | || || || | |||| || | ||| |||||||| || || |||

RESULT   3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENTER, John C.
;   TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;   TITLE OF INVENTION: TUBERCULOSIS
;   FILE REFERENCE: 24366-20007.00
;   CURRENT APPLICATION NUMBER: US/09/103,840A
;   CURRENT FILING DATE: 1998-06-24
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match   3.3%;   Score 49.8; DB 4; Length 4411529;
Best Local Similarity 48.2%; Pred. No. 0.089;
Matches 172;   Conservative   0; Mismatches 182; Indels   3; Gaps   1;

Qy   695 gcaccaggctcctccgcctgacgcgccttcgcgcctgtgcgcatctcgcaggag 754
          || | || | || |:|| |||||| | ||||:| | ||| | |||||
Db 337017 gcgcgcggcgccgaacagcgccgtgcccccggcgcccgcgtcggcgcgagcgtgcc 337076

Qy   755 cgcctctgccgtgccgtcgcctttgccgtcaccgccgccagcctcagcccacctcggcc 814
          |||| | | :| ||||: | | :|||||| | | ||| ||| ||| || |
Db 337077 cgccccgccggtgccgcggcgcggcgcgaagagcaatcgttcccccggtcccgccattcgc 337136

Qy   815 tcattgccctgggcacttctgggcctgtgttggc---caggtgacagtgccacggcc 871
          | | | || || || | |||| || | ||| |||||||| || || |||
Db 337137 gcaaacccgccgccgtcccgccgcccgccgccggtctggcgaacagcccaccggtaccaccggc 337196

Qy   872 tcctgcagccgtccccagctcggtctcgagggccaggcagtggacagtgggcgg 931
          ||| ||| | || | | | | |||||||| |||| | |||||| | | |
Db 337197 tccggcggtgccggcgccgccgccaccgataaagtttgggagagggcggcggcacgacctgcgcg 337256

Qy   932 ggtcagacaccgagggcctccgatgggcggaagggctgggcggcaccgacctgcgggggc 991
          ||||| || || || ||||| ||| || | ||||| | |||||||| | |||
Db 337257 tgcggcaccgaggacaagccggcgtcaccgccgtcggcgcgccgccgccgccgtcggcgggccc 337316

Qy   992 cggacctggccaactgcgtagcactcgcctcgctcgagcaccgatcgcgcgctgcc 1048
          |||| |||| | ||| | || | | ||||| | ||| ||| | | | |||
Db 337317 gccaaacccgcgctgcgccggtgcgccgccgagcgatcagcaaggcggctccgcc 337373

RESULT   4
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202

; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
-105-537-1

```
Query Match           3.0%;   Score 46; DB 4;  Length 15872;
Best Local Similarity 44.3%;  Pred. No. 0.13;
Matches 234;  Conservative   0;  Mismatches  290;  Indels   4;  Gaps   1;

Qy    557 cctgcggccgccactcccgtctccccgtgctcggctcagcatggccgccgtcacaaactc 616
          ||| || ||| ||||| |  ||||| |   ||||||| ||||| ||| || ||  |   |
Db   3016 cctgggcgacctgctcaccggcggcagcggcgagaccggaccggatcggcgacgggataccgcc 3075

Qy    617 tgtagccctgacccgggccagccaacaacagtgaaggcagcgaccatgactacctgccc 676
          ||  |  ||| ||| || |||  ||||| |||    ||| | |||||| ||
Db   3076 cggaccccggcggacaccacccggcacccgccgagcctgcgatcatggcatggcctgccgcta 3135

Qy    677 ttggtgcggctgcaggaggcaccaggctctt----cgcctggacggcgccctctgcggc 732
          |||| |  |||  |  |     |||  |  ||    |||  | ||| ||||| |||
Db   3136 cccggcgggcgtcacctcccgaggacctgtggcggctcgtcgcgggggcgcgacgc 3195

Qy    733 cgctgcgcatctgcaggagagccgcctgctgccgctgcctgccttgcgtgcaccggcgc 792
          || |  |   ||| |  |  | |||| |  ||||| |||  ||||| ||||| |  |
Db   3196 cgtctcggggctgcccaccgacccggctggacggcagaggacctcttcgacgcgaccccga 3255

Qy    793 cagcctcagcccaccctcggcctcattgccctgggcacttctggccctggtgttggcca 852
          | |  | ||| | |||| |  | | ||||   |   | ||||| |
Db   3256 ccgcagcggcaaagagctcggtccggagggcggcggattcctgcacgacgccccgttcga 3315

Qy    853 gggtgacagtgccacggcctcctgcagacaccgaggcctccgatggggctcgagggccagg 912
          ||||| |||  ||||||  | ||| | || || | ||   |||||   ||||| |||| ||
Db   3316 cgccggcttcttcgggatatcgccccgcgaggcctgcatggacgccgcagcagcagcggct 3375

Qy    913 ccaggtggacagtggcgggggtcagacacggcctccgatggggcggaaggggctggg 972
          |  ||| |    | || | || | | |   |  || || || |   | ||| |   |||
Db   3376 gctcctggagacggacggcatggaggcggcgtggagcggcgcgcagggctcgacccggaaggcctcaa 3435

Qy    973 cggcaccgacctgcggggccggacctgggccactccggctagcactccggcctggtggagca 1032
          ||||   |  |  | |||  ||||  ||||| | || ||  ||  |  ||| ||  || ||
Db   3436 gggcgccgacggcggcggcgtcttcgtcggccacctcggccccgcccgcccgccgccgatgca 3495

Qy   1033 ccgatgcgccgctgccttcgacgagtgggaactgacagcggccaaggc 1080
          ||||||   | ||| |  ||  || ||| ||| |  |  | ||| || |||
Db   3496 cgacggcgccgagggcgtcgagggcacctcctgaccgggaccacgcc 3543
```

RESULT 5
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```
Query Match           3.0%;  Score 45.6; DB 4;  Length 4403765;
Best Local Similarity 46.3%; Pred. No. 0.82;
Matches 230; Conservative   0;  Mismatches  262;  Indels   5;  Gaps   2;

Qy    560 cgccgcccactcccgtctccccgtcggctcagcatggccgccgtcacaaactctgt 619
          || || ||| ||| |   ||  |||| ||  |||   ||||| || ||  | |
Db 3158094 CGGCCCCGGCCGCCACCGAGGCCGGGGCTCCGCCGGCCACCGCGTCACCGCCATTGC 3158035

Qy    620 agcctgacccggccaggccaacaacagtgaaggcagcgaccatgactacctgccctg 679
          | ||||| || || |  |  |    |  |||   |   | ||||| |||  |  |
Db 3158034 CGCCAGTACC--GCCCGCCCTCGTCACCGCCGCCGGGAGTTGGTGGACGCCCCGCCGGG 3157977

Qy    680 gtgcggctgcaggaggcaccaggctcctgggaccgggccttctgccgcctgtg 739
          ||||||| |||| || || ||  ||||| ||||  |||| |||||| ||| |
Db 3157976 AATAGCTTCGCCTTCGGCCCCGGCCGGACCGCCCTCACCCGCCGCCACCGCGCCACCGCCGCC 3157917

Qy    740 cgcatctcgcaggagcgcctcgcctgcctgccctttgccgtgcaccggcgcagcctc 799
          ||| |   ||| |  | ||  ||| ||  ||| | |||| |
Db 3157916 CGCACCGCCGAGGCCGCAGCCGCCGGCGGCGCCGCTGCCAATCAGCAGCCCGCCTGC 3157857

Qy    800 agcccacctcggcctcattgcctgggcacttctgggccctgtgttggcaggtgac 859
          || || ||| ||||  |||  | | ||||||  |  || ||||  || |||||
Db 3157856 ACCCCCGTACCACTAGACCTCCGGCACCCCCGTCCGGCGCTGCC 3157797

Qy    860 agtgccacggcctcctgcacgtctcgggctcggctctgagggggccaggtg 919
          | ||||||||  | ||   | |  | ||| | || || ||| ||| | |||
Db 3157796 GCCGTCTCCGCGTCCGCCATTGCCCCGGTTACGG---TGGCATTGTCGCCGAGGCT 3157740

Qy    920 gacagtgggggtcagacaccgaggcctccgatggggcggaaggggtgggcggcacc 979
          |||| ||||  ||| | |  | ||||| ||||   || |  ||| | || |  |||
Db 3157739 GTTGGCGATGTCGGCTTGGCTTTGCCGGCCGCCGGGGGCCACCGCCGCCACC 3157680

Qy    980 gacctgcgggccggacctgggccactgccgctagcaactgccgctggctgccgctgagcaccgatgc 1039
          |||| ||||  || || ||||| |  |  || |  | ||||  || || || || | ||||
Db 3157679 GGTACCGCCGTTGCCACCGGCACCGGCGGTGCCCGGTGCCGGCGTTGCCGGCCAGCAGGCGCCACG 3157620

Qy   1040 gccgctgcttcgacga 1056
          || ||  |||||| ||
Db 3157619 GCCCGCGTCACCGCCGA 3157603
```

RESULT 6
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae

US-09-105-537-32

```
Qy   831 ttctgggccctggtgttggccaggtgacagtgccaccgcgcctcctcctgcagcccgtccccca 890
         || |||||||  |  |  |     ||| |   |||  || |||||    ||||| |||||||
Db 24695 gcatgtacccgatccggcgctgatg-----gggccaggccagtggacaagtggcggggctcagacaccgcgaccg 24754
Qy   891 gctcgggtctgag-----gggccaggccagtggacaagtggcggggctcagacaccgcgt 944
         ||  | ||                |||||  || |    |||  | | || ||  | | |
Db 24755 gccccggcgtcacgcacctcgccccgaccgggtcatgggcctgctctccgggcccgacctgggcca 24814
Qy   945 aggctccgatggggcggaagggctgggcgacgcgcggaccctcgggggccgacctgggcca 1004
         |   ||||      |     ||| || || || |||||| || ||    | ||| ||| |
Db 24815 acgccccggtcgtcgtggcggacgcgcggaccctcgcgggatgccgctgccttcgacgagtgggacg 24874
Qy  1005 ctgccgtagcactcgcctggctggagcaccgatgcgccgcgctgccttcgacgagtgggaac 1064
         ||| || || ||||| || | ||  |||| ||  | |  |||||||||||||||||||||
Db 24875 tcgcccaggggcgcctccgtccgtgcgtgttcctgacggcgtctacgccctgacggccttgacc 24934
Qy  1065 tgacagcggccaaggctgattgctggctgcgggccgagcgcctccggtccactccgcgcccggtggcgt 1124
         ||  | ||  ||| || ||  |  || || ||||||  |   | ||||||||||||||||||||||||
Db 24935 tggcggacgtcaagcccgcgagcgcctcctggtccactccgcgcccggtggcgtgggca 24994
Qy  1125 tgccgccctcaaggccgc 1143
         |||||||  | || | ||
Db 24995 tgccgccgtgcagctcgc 25013
```

RESULT 8
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```
Query Match        2.9%;  Score 44.2;  DB 3;  Length 38506;
Best Local Similarity  45.6%;  Pred. No. 0.45;
Matches 200;  Conservative 0;  Mismatches 233;  Indels 6;  Gaps 1;

Qy   711 gcctggacggcgcccttctgcgcgtgtgcgcatctcgcaggagagccgcctctgccgtgcct 770
         |||| | |||||||||| || |  |  |  || ||| |||||||  || ||||| ||||  |
Db 22717 gcctcacggcggcgcccggcgacgccgagaccctcgccccggagccgctcggccccgggac 22776
Qy   771 cgccctttgccgtgcaccgcgatccggcaccggtctcaacttccgcgacgtcctgggcac 830
         |||   | ||  |  |  |  ||  ||||||||||||||||||||||||||||| ||  ||
Db 22777 aggtccgatcgcgatccggcaccggtctcaacttccgcgacgtcctgatcgccctcg 22836
```

US-09-105-537-32

```
Query Match        2.9%;  Score 44.2;  DB 4;  Length 11220;
Best Local Similarity  45.6%;  Pred. No. 0.32;
Matches 200;  Conservative 0;  Mismatches 233;  Indels 6;  Gaps 1;

Qy   711 gcctggacggcgcccttctgcgcgtgtgcgcatctcgcaggagagcgcctctgccgtgcct 770
         |||| | |||||||||| || |  |  |  || ||| |||||||  ||||||| ||||  |
Db  8888 gcctcacggcggcgcccggcgacgccgagaccctcgccccggagccgctcggccccgggac 8947
Qy   771 cgccctttgccgtgcaccgcgatccggcaccggtctcaacttccgcgacgtcctgggcac 830
         |||   | ||  |  |  |  ||  ||||||||||||||||||||||||||||| ||  ||
Db  8948 aggtccgatcgcgatccggcaccggtctcaacttccgcgacgtcctgagcgccctcg 9007
Qy   831 ttctgggccctggtgttggccaggtgacagtgccaccgcgcctcctcctgcagcccgtccccca 890
         || |||||||  |  |  |     ||| |   |||  || |||||    ||||| ||||||
Db  9008 gcatgtacccgatccggcgctgatg-----gggccaggccagtggacaagtggcggggctcagacaccg 9067
Qy   891 gctcgggtctgag-----gggccaggccagtggacaagtggcggggctcagacaccg 944
         ||  | ||                |||||  || |    |||  | | || ||  | |
Db  9068 gccccggcgtcacgcacctcgccccgaccgggtcatggggcctgctctccgggcgt 9127
Qy   945 aggctccgatggggcggaagggctgggcgacgcgcggaccctcgggggccgacctgggcca 1004
         |   ||||      |     ||| || || || |||||| |||||    | ||| ||| |
Db  9128 acgccccggtcgtcgtggcggacgcgcggaccgtcgcgggatgccgcggtggcgtgggcagt 9187
Qy  1005 ctgccgtagcactcgcctggctggagcaccgatgcgccgcgctgccttcgacgagtgggaac 1064
         ||| || || ||||| || | ||  |||| ||  | |  |||||||  | | || |||  |
Db  9188 tcgcccaggggcgcctccgtccgtggtgttcctgacggccgctacgccctgccgcacc 9247
Qy  1065 tgacagcggccaaggccgattgctggctgcgggccccagcacttgcctgacggccttgacc 1124
         ||  | ||  ||| || || | | || || || ||||  |   | | |||||||||||||
Db  9248 tggcggacgtcaagcccggcagcccggcgcctcctggtcactccgcgcccggtggcgtgggca 9307
Qy  1125 tgccgccctcaaggccgc 1143
         |||||||  | || | ||
Db  9308 tgccgccgtgcagctcgc 9326
```

RESULT 7
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

```
Query Match        2.9%;  Score 44.2;  DB 4;  Length 36778;
Best Local Similarity  45.6%;  Pred. No. 0.44;
Matches 200;  Conservative 0;  Mismatches 233;  Indels 6;  Gaps 1;

Qy   711 gcctggacggcgcccttctgcgcgtgtgcgcatctcgcaggagagcgcctctgccgtgcct 770
         |||| | |||||||||| || |  |  |  || ||| |||||||  ||||||| ||||  |
Db 24575 gcctcacggcggcgcccggcgacgccgagaccctcgccccggagccgctcggccccgggac 24634
Qy   771 cgccctttgccgtgcaccgcgatccggcaccggtctcaacttccgcgacgtcctgggcac 830
         |||   | ||  |  |  |  ||  ||||||||||||||||||||||||||||| ||  ||
Db 24635 aggtccgatcgcgatccggcaccggtctcaacttccgcgacgtcctgatcgccctcg 24694
```

```
Qy    831  ttctggcccctggtgttggccagggtgacagtgccacggccgcctcctgcagccgcgtccccca  890
           |||||||||||||| |||||||||||||||||  | ||||||||||||||||||||||||||||
Db  22837  gcatgtaccccgatccggccgctgatggcaccgagggagccggcgtggtcaccgcaccg       22896

Qy    891  gctcggcttgag------ggcaggccaggtggacagtggacagtggggctcagacaccg        944
           ||| ||||   |||       |||||||||||||||||||||||||||||||||||||||
Db  22897  gccccgcggtcacgcacctcgccccggcgaccgggtcatgggcctgctcctccggcgcggt    22956

Qy    945  aggcctccgatgggcggaagggctgggcggcaccgacctcggggccggacctgggcca       1004
           | ||| |||||||||||||||||||| |||||||||||||||||||||  | ||| |||
Db  22957  acgccccggtcgtcgtggcggacggcgcggaccgtcgcggggatgcggagcgggggttggacgt  23016

Qy   1005  ctgccgtagcactcgcctgcctggagcacggcaccgatgcgccgcctcgacgagtgggaac   1064
           | |||| |||||  ||||||||||||||| |||| | ||||||||||||| ||| ||||
Db  23017  tcgcccaggggcgcctccgtgccggtggtgttcctgacggcctctacgccctgcgcgacc    23076

Qy   1065  tgacagcggccaaggctgatgattgctgctgcgggggccaacacttgctgctgacccgaccgcctgacc  1124
           |||  ||||| |||   ||| | |||||| | |||||||||||||| |||  ||| |||  ||| |||
Db  23077  tggcggacgtcaagcccggcgagcgcctcctggtccactccgccgccggttcctgacgccctggcggca   23136

Qy   1125  tggccgccctcaaggccgc  1143
           ||||||||| || |||
Db  23137  tggccgccgtgcagctcgc  23155
```

```
RESULT   9
US-08-343-428-1
; Sequence 1, Application US/08343428
; Patent No. 5665586
; GENERAL INFORMATION:
;  APPLICANT:  Nakamura, Etsuo
;  APPLICANT:  Tsuzuki, Hiroshige
;  APPLICANT:  Kitadokoro, Kengo
;  APPLICANT:  Shin, Nasaru
;  APPLICANT:  Teraoka, Hiroshi
;  TITLE OF INVENTION:  No. 5665586el Protease
;  NUMBER OF SEQUENCES:  4
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  James F. Haley, Jr., Fish & Neave
;   STREET:  1251 Avenue of the Americas
;   CITY:  New York
;   STATE:  New York
;   COUNTRY:  USA
;   ZIP:  10020-1104
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Diskette, 3.50 inch, 720Kb storage
;   COMPUTER:  IBM Compatible
;   OPERATING SYSTEM:  MS Dos 5.0
;   SOFTWARE:  WordPerfect 5.0
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/343,428
;   FILING DATE:  18-NOV-1994
;   CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  PCT/J93/00592
;   FILING DATE:  30-APR-1993
;   APPLICATION NUMBER:  JAPAN 4-126511
;   FILING DATE:  19-MAY-1992
;  ATTORNEY/AGENT INFORMATION:
;   NAME:  Haley, Jr., James F.
;   REGISTRATION NUMBER:  27794
;   REFERENCE/DOCKET NUMBER:  SHGN-7
;  TELECOMUNICATION INFORMATION:
;   TELEPHONE:  (212)596-9000
;   TELEFAX:  (212)596-9090
; INFORMATION FOR SEQ ID NO:  1:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH:  2064
;   TYPE:  nucleic acid
;   STRANDEDNESS:  double
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  Genomic DNA
;   ORIGINAL SOURCE:
```

```
;    ORGANISM:  Streptomyces fradiae
;  FEATURE:
;   NAME/KEY:  -35 signal
;   LOCATION:  359..364
;   IDENTIFICATION METHOD:  by experiment
;   NAME/KEY:  -10 signal
;   LOCATION:  378..383
;   IDENTIFICATION METHOD:  by experiment
;   NAME/KEY:  CDS
;   LOCATION:  435..1505
;   IDENTIFICATION METHOD:  by experiment
;   NAME/KEY:  sig peptide
;   LOCATION:  435..944
;   IDENTIFICATION METHOD:  by experiment
US-08-343-428-1
```

```
Query Match             2.9%;  Score 44;  DB 1;  Length 2064;
Best Local Similarity  43.8%;  Pred. No. 0.22;
Matches  191;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;
```

```
Qy    700  aggctccttcgcctgagcgccctctgcgccgtgtcgcatccgcaggagcgcct  759
           |||| ||| |  || |||   ||| |||||  |  ||   |||||| ||||||||
Db    440  ACGCACCACCCGCGCCGCACCGGTCTGTCCGCCCGCTCCTCGGCGCCTCCTCGGCCT  499

Qy    760  ctgccgtgcctgcgcctttgccgtgcaccgcgccagcctcagcccacctcggcctcatt  819
           | ||  |||  ||  ||| ||| ||  ||| ||||||||||||| ||||| |||||||||
Db    500  CGGCGGCGCCGCCCGGCGGAGCGGAGGCGCCCCCAGAGGCGCCGCCCCCACCCGGCCTCCGA  559

Qy    820  gccctggcactctggcctcggttggccaggtgacagtgccacggcctcctcgcag  879
           | |    |||||| |||| ||||  ||   ||  ||  ||| |||   ||||| |||
Db    560  CTCCGCCGGCAGCCCTCCACGCGCCTCGACGGCGCGCGTCGAGCGGACGGACGACAG  619

Qy    880  cccgtccccagctcgggctctgagggggccaggccagtgacagtgggcgggctcaga  939
           |||  ||   |||||||||  |||| ||  ||| |||   ||| || ||   |||||
Db    620  CGCCGGCACCTACGTGACGCGGCACCGGCGAACTCGTCGTCACCGTCACCACCGAGGC  679

Qy    940  caccgggcgctccgatgggcggaagggctgggcggcacccgaccctgcggggccggaccctg  999
           || |||||||  | |  |||  ||| |||||||   ||   || |  |||||   ||||| |
Db    680  CGCCGCCGCCAAGGTCCGACGGCGGCGGCGGCGGCCCCGGGGGTCCAGCGGGCGGCGGGCCGC  739

Qy   1000  ggccactgccgtagcactcgcctggctggagcaccgatgcgccgccttcgacgagtg  1059
           || |||| ||   ||  |   |||||| | ||| |||  |||||  ||   |||  ||
Db    740  CGAGCTCGACGGCGCCATGGCCCCTGGAGCACGGCGGCCGCAGATCCCGGCCACCTCGTG  799

Qy   1060  ggaactgacagcggccaagcggccaagctgattgctgctgcggcccagcacttgctgacggcct  1119
           |||   | |  | |||  | |||  |||  |    |||| ||||||| || ||  || |  ||||
Db    800  GGGGCTGGACCCGGCGACCAACCGGGAGATCGCCGTGGGAGGGCCGGACTCCTCCTCGCGCCCCG  859

Qy   1120  tgacctggccgccctc  1135
           |||||||| |  ||
Db    860  CGACCTGGCCCGGCTC  875
```

```
RESULT  10
US-08-537-002A-5
; Sequence 5, Application US/08537002A
; Patent No. 5773282
; GENERAL INFORMATION:
;  APPLICANT:  TSUSAKI, Keiji
;  APPLICANT:  KUBOTA, Michio
;  APPLICANT:  SUGIMOTO, Toshiyuki
;  TITLE OF INVENTION:  RECOMBINANT THERMOSTABLE ENZYME FOR
;  TITLE OF INVENTION:  CONVERTING MALTOSE INTO TREHALOSE
;  NUMBER OF SEQUENCES:  17
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  BROWDY AND NEIMARK
;   STREET:  419 Seventh Street, N.W., Suite 300
;   CITY:  Washington
;   STATE:  D.C.
;   COUNTRY:  USA
;   ZIP:  20004
;  COMPUTER READABLE FORM:
```

```
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/537,002A
      FILING DATE: 29-SEP-1995
      CLASSIFICATION: 435
PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 260984/1994
      FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP No. 5773282 yet received
      FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
      NAME: BROWDY, Roger L.
      REGISTRATION NUMBER: 25,618
      REFERENCE/DOCKET NUMBER: TSUSAKI-1
TELECOMMUNICATION INFORMATION:
      TELEPHONE: 202-628-5197
      TELEFAX: 202-737-3528
      TELEX: 248633
ORMATION FOR SEQ ID NO:   5:
EQUENCE CHARACTERISTICS:
      LENGTH: 3600 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS: DOUBLE
      TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
      ORGANISM: Thermus aquaticus
      INDIVIDUAL ISOLATE: ATCC 33923
FEATURE:
      NAME/KEY: 5'UTR
      LOCATION: 1..540
      IDENTIFICATION METHOD:      E
      NAME/KEY: mat peptide
      LOCATION: 541.3429
      IDENTIFICATION METHOD:      S
      NAME/KEY: 3'UTR
      LOCATION: 3430..3600
      IDENTIFICATION METHOD:      E
US-08-537-002A-5
```

```
Query Match          2.9%;   Score 43.6;  DB 1;  Length 3600;
Best Local Similarity  44.4%;   Pred No. 0.31;
Matches 175;  Conservative   0;  Mismatches 219;  Indels   0;  Gaps     0;

Qy   659 gaccatgactacctgccttggtcggctgcaggaggaggcacaggctccttccgcctggac  718
          | |||   || ||  |  ||  | ||  |||||    || ||    |  |||| |
Db  3070 GGCCGCGGCCTCGGGGCCTTCCTGGAGCTGGAGGGGGAGGTGTACCTCGTGGCCCTGGGC 3129

Qy   719 gcgcccttctgccgcgctgtgcgcatctcgcaggagcgcctcgccgtgcctcgccctt   778
          |||       || |||  ||  || | ||   ||    ||||| | |  |  |  | |
Db   130 GCGGAAAAGCCGGGCACGGTGGAGGAGGACCTGGCCCGCCTACGACGTGGAGCGG     3189

Qy   779 gccgtgcaccgcggcagcctcagcccacctcggcctcattgccctgggcacttctggc   838
          | || ||   ||| || | |         ||   | |     |   |   ||  ||
Db  3190 GCCCTGCACCTCGCCCTCCAGGCCCTGGAGGCCGAGCTTTGCCGCCTTTGCCGAGGAGTG 3249

Qy   839 cctggtgttggccaggtgacagtgcacgccgcctcctgcagccgtccccagctcgggc   898
          |             | |      |  |||  |||      ||  ||  ||| |  |||
Db  3250 GCCGACCACCTCACGCGCCGCCTTCCTCCAAGCCTACCGCTCCGCCCTCCCCGAGGAGGCC 3309

Qy   899 tctgaggggccagccaggtggacagtggcgggctcagacaccgaggcctccgatggg    958
          | ||| |  |  |||  ||  |||  |   || |  || | |  | |||  |   |
Db  3310 CTGGAGGAGGCGGGCTGACGCGGCCACATGGCCGGCACGGTGGCGGGAGCACCTCCACCGG 3369

Qy   959 gcggaagggctgggcggcaccgacctcggggccggaacctggccactgcgtagcactc 1018
          || ||||  ||| || || |   |||     ||| |  || |||||  |  || |   |
Db  3370 GAGGAAAGCCCCGGCCATCCACGAGCGCTCGCCAGCCAAGGCCCGAAAGCC     3429

Qy  1019 gcctggctggagcaccgatgcgccgctgccttcg 1052
          ||  ||    | ||  |  ||||| |  |  | ||
Db  3430 TAGGCGCCCGGTAGCCCTTCAGCCCGGGGCCACG 3463

Search completed: September 19, 2002, 07:37:59
Job time: 24049 sec
```

OM of: US-09-786-136-4 to: SPTREMBL_19:*   out_format : pfs

Date: Sep 19, 2002 5:45 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=frame+_n2p.model -DEV-x1h
  -Q=/cgn2_1/USPTO_spool/US09786136/runat_16092002_141244_19221/app-query.fasta_1.1592
  -DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
  -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
  -ANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
  -X_MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
  -RM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
  -USER=US09786136 @CGN_1_100 -NCPU=6 -ICPU=3 -LONGLOG
  -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
  Query: US-09-786-136-4
  Query length: 1511
  Database: SPTREMBL_19:*
  Database sequences: 562222
  Database length: 172994929
  Search time (sec): 94.020000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
sp_human:Q9BVH8	+	1778.50	2027.29	1.4e-105	1001	! Q9bvh8 homo sapiens (human). h
sp_mammal:Q9GMT9	+	1625.50	1854.11	7.7e-96	803	! Q9gmt9 macaca fascicularis (cra
sp_mammal:Q9GKV4	+	1277.00	1464.88	1.1e-73	260	! Q9gkv4 macaca fascicularis (cra
sp_rodent:Q9CUE8	+	359.50	402.55	3.7e-15	1209	! Q9cue8 mus musculus (mouse). ba
sp_human:Q9BX42	+	240.50	271.89	1.1e-07	628	! Q9bx42 homo sapiens (human). ba
sp_plant:Q9FPQ6	+	235.00	266.63	3.0e-07	555	! Q9fpq6 chlamydomonas reinhardti
sp_bacteria:Q9XDH2	+	233.00	261.68	4.1e-07	763	! Q9xdh2 mycobacterium tuberculos
sp_plant:Q9SPM0	+	226.50	249.71	1.1e-06	1315	! Q9spm0 zea mays (maize). exten
sp_virus:O41935	+	224.00	253.61	1.5e-06	585	! O41935 murid herpesvirus 4. hyp
sp_fungi:O42854	+	218.00	238.29	3.9e-06	1611	! O42854 schizosaccharomyces pom
sp_virus:P89459	+	218.00	232.77	4.1e-06	3122	! P89459 herpes simplex virus (t
sp_bacteria:O68872	+	217.50	246.81	3.9e-06	542	! O68872 myxococcus xanthus. hypo
sp_invertebrate:Q9VZC2	+	215.00	245.57	5.5e-06	446	! Q9vzc2 drosophila melanogast
sp_invertebrate:Q9VPG1	+	215.00	231.95	6.2e-06	2284	! Q9vpg1 drosophila melanogast
human:Q9H6K5	+	214.00	241.74	6.5e-06	616	! Q9h6k5 homo sapiens (human). co
fungi:O36027	+	210.50	238.32	1.1e-05	574	! O36027 schizosaccharomyces pomb
vertebrate:O57580	+	210.50	232.52	1.1e-05	1151	! O57580 gallus gallus (chicken)
sp_virus:O65553	+	210.00	223.29	1.3e-05	3247	! O65553 bovine herpesvirus 1. u
sp_plant:Q41805	+	209.50	231.11	1.3e-05	1188	! Q41805 zea mays (maize). exten
sp_virus:O41972	-	207.00	232.59	1.8e-05	706	! O41972 murid herpesvirus 4. hyp
sp_virus:O41973	+	207.00	232.35	1.8e-05	727	! O41973 murid herpesvirus 4. hyp
sp_virus:O41971	-	207.00	231.82	1.8e-05	774	! O41971 murid herpesvirus 4. hyp
sp_virus:Q69146	-	204.00	233.16	2.8e-05	437	! Q69146 human herpesvirus 4. (eps
sp_mammal:Q95JD1	+	203.00	229.86	3.3e-05	566	! Q95jd1 sus scrofa (pig). basic
sp_human:O75851	+	201.50	212.15	4.3e-05	4123	! O75851 homo sapiens (human). w
sp_invertebrate:Q9BIT7	-	201.50	217.81	4.4e-05	1953	! Q9bit7 nephila madagascarien
sp_virus:Q9Q5L3	+	201.00	227.02	4.4e-05	605	! Q9q5l3 cercopithicine herpesvir
sp_virus:Q9IPQ9	-	201.00	226.83	4.4e-05	619	! Q9ipq9 cynomolgus epstein-barr
sp_mammal:Q95JD0	+	199.50	226.71	5.4e-05	511	! Q95jd0 sus scrofa (pig). basic
sp_invertebrate:Q93107	+	197.50	221.19	7.4e-05	753	! Q93107 acanthamoeba castellan
sp_bacteria:Q9RX57	+	197.50	220.29	7.5e-05	839	! Q9rx57 deinococcus radiodurans
sp_human:Q9ULT1	+	197.50	217.85	7.6e-05	1124	! Q9ult1 homo sapiens (human). k
sp_mammal:Q95JC9	+	196.50	220.94	8.5e-05	676	! Q95jc9 sus scrofa (pig). basic
sp_invertebrate:Q9N968	+	196.00	218.79	9.3e-05	817	! Q9n968 leishmania major. hypo
sp_human:Q9UDT9	+	195.50	216.08	0.0001	1056	! Q9udt9 homo sapiens (human). w
sp_bacteria:Q9F2N5	+	194.00	215.80	0.0001	889	! Q9f2n5 streptomyces coelicolor.
sp_human:Q9Y5L9	+	193.50	205.16	0.0001	2971	! Q9y5l9 homo sapiens (human). t
sp_human:Q9P206	+	193.00	213.31	0.0001	1044	! Q9p206 homo sapiens (human). k
sp_rodent:Q9R172	+	193.00	206.66	0.0002	2319	! Q9r172 rattus norvegicus (rat)
sp_fungi:Q9P6T1	+	192.50	207.35	0.0002	1992	! Q9p6t1 neurospora crassa. hypo

sp_plant:Q9FLQ7   + 192.00 210.41 0.0002   1289 ! Q9flq7 arabidopsis thaliana
sp_invertebrate:Q9NE65   + 191.50 207.16 0.0002   1778 ! Q9ne65 leishmania major.
sp_rodent:Q63123   - 191.00 208.47 0.0002   1419 ! Q63123 rattus norvegicus (r
sp_invertebrate:Q9BIU7   - 189.50 213.29 0.0002   648 ! Q9biu7 argiope trifasciata
sp_rodent:P70433   + 189.50 211.51 0.0002   802 ! P70433 mus musculus (mouse).

seq_name: sp_human:Q9BVH8

seq_documentation_block:
  ID Q9BVH8   PRELIMINARY;   PRT;   1001 AA.
  AC Q9BVH8;
  DT 01-JUN-2001 (TrEMBLrel. 17, Created)
  DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
  DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
  DE HYPOTHETICAL 106.2 KDA PROTEIN (FRAGMENT).
  OS Homo sapiens (Human).
  OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  OX NCBI_TaxID=9606;
  RN [1]
  RP SEQUENCE FROM N.A.
  RC TISSUE=EYE, RETINOBLASTOMA;
  RA Strausberg R.;
  RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
  DR EMBL; BC001200; AAH01200.1; -.
  DR InterPro; IPR002035; vWFA.
  DR SMART; SM00327; VWA; 1.
  KW Hypothetical protein.
  FT NON_TER   1   1
  SQ SEQUENCE   1001 AA;   106194 MW; 9DFD75263D040E43 CRC64;

alignment_scores:
      Quality: 1778.50   Length: 400
      Ratio: 5.111   Gaps: 3
  Percent Similarity: 87.000   Percent Identity: 85.250

alignment_block:
  US-09-786-136-4 x Q9BVH8

Align seg 1/1 to: Q9BVH8   from: 1 to: 1001

   86 TGGGGAGTCTGAGCTGGTTCTGGCCTAGCAAGGCAATCTACTCTGTTAA 135
      |||||||||||||||||||||||||||||||||||||||||||
  623 TrpGlyProGlyAspGlySerGlnProProSer............... 633

  136 AGTCCGATAGAGGGAGAAGCTGGGAGGAACACTGCAATTTTTCATGAGC 185
      ||||||||||||||||||||||||||||||||
  634 ....ProP 635

  186 CCTCTGTAGAGGGTGGGTGG...TGGGGGGTCCACAGACTGCTCCTTGGT 232
      ||||||||||:::||| ::|||||||||||
  635 roValArgGluAlaAlaTrpAspGlnAlaLeuHisArgLeuThrAlaAla 651

  233 TCCACA.. 238
      |||:::
  652 SerValArgAspAsnGluGlnLeuAlaLeuArgGlyGlyAlaGluuTh 668

  239 .........GGCCATGCCCGGAGGTGCTGGCTTCAGCCCCTTCAGA 275
      |||||||||||||||||||||||||||||||||
  668 rThrAlaAspArgGlyHisAlaArgArgCysTrpLeuArgAlaLeuGlnT 685

  276 CAAGTAAGGTCAGCTCGCCCCTCCTGCTTCACTTGCCCTGTAGCTGTG 325
      |||||||||||||||||||||||||||||||||||
  685 hrSerLysValSerSerAlaProSerCysPheThrCysProValAlaVal 701

  326 GATGCTACTACTAGGGAGGTCCTGCCTGGGGCCCTGCCAGGTGTGCAGCTC 375
      ||||||||||||||||||||||||||||||||||
  702 AspAlaThrThrArgGluValLeuProGlyValAlaLeuGlnValCysSerSe 718

  376 AGAGCCCGCTGAGCCCCAGGAACCCCTGCTTCCTCTCACAGCCATCTAG 425
      ||||||||||||||||||||||||||||||||||
  718 rGluProAlaGluProProGlyThrProProAlaSerHisSerHisLeuA 735

```
DE   HYPOTHETICAL 84.7 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047829; BAB12255.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   803 AA;  84655 MW;  724FC68DCC2D1D92 CRC64;
```

```
alignment_scores:
              Quality:  1625.50                   Length:      400
                Ratio:    4.926                      Gaps:        4
  Percent Similarity:   82.500        Percent Identity:   79.500
```

```
alignment_block:
US-09-786-136-4 x Q9GMT9       ..

Align seg 1/1 to: Q9GMT9  from:  1  to:  803

 86 TGGGGAGTCTGAGCTGGTTCTGGGCCTAGCAAGGCAATCTACTCTGTTAA 135
    |||||   |||| |||        ||||||        ::
433 TrpGlyProGlyAspGlySerGlnProProSer.................. 443

136 AGTCGATAGAGGGAGAAGCTGGGAGAACACTGCAATTTTTCTATGAGC    185
                                                    |
444 ..................................................ProP 445

186 CCTCTGTAGAGGGTGGGTGG...TGGGGGCTCCACAGACTGCTCCTTGGT 232
    ||    |||:|||||          ::|||||||||||  |:::
445 roValArgGluAlaAlaTrpAspGlnAlaLeuHisArgLeuThrAlaAla  461

233 TCCACA............................................. 238
    |||:::
462 SerValValArgAspAsnGluGlnLeuAlaLeuArgGlyGlyAlaGluTh 478

239 ..........GGCCATGCCCGGAGGTGCTGGCTTCAGCCCTTCAGA     275
              |||||||||||||||||||||  |||||||||||||
478 rThrAlaAspArgGlyHisAlaArgArgCysTrpLeuArgAlaLeuGlnT 495

276 CAAGTAAGGTCAGCTCTGCCCCCTCCGTTCACTTGCCCTGTAGCTGTG   325
    ||||||||||||||||||||||||||||||||||||||||||||||||
495 hrSerLysValSerSerAlaProSerCysPheThrCysProValAlaVal 511

326 GATGCTACTACTAGGAGGAGGTCTGCCTGGGCCCTGCAGTGTGCAGCTC  375
    ||||||||||||||||||||||||  ::|||||||||||||||||||||
512 AspAlaThrThrArgArgGluValLeuProGlyValGlnValCysSerSe 528

376 AGAGCCCGTGAGCCCCAGGAACCCCTCCTGCCTCTCACAGCCATCTAG  425
    ||||||||||||||||||||||||||||||:::||||:::::|||
528 rGluProValGluProGlyThrProProAlaAlaHisSerArgLeuA   545

426 ATGCAGCTCCTCTGCCCACTGTGTCTACTCTAAAGGACTTCAGAGAGGC  475
    |||||||||||||||||||||||||||||||||||||||||||||||||
545 spAlaAlaProLeuProThrValTyrSerLys...............   556

476 TCTCCAGCAGGGCGCCTGGGACTCGGACCAAAATGGAACTCCAAGCGTGC 525
    ||||||||||||||||||||||||||||||||||||||||||||||:  |
557 ........GlyAlaTrpAspSerArgAspArgAsnGlyAsnSerLysCysAl 570

526 TTTGGGGACCCTGCCACTCCCACGGAAGGTCCTGCCGCCCACCTCCCC   575
    |||||||||||||||||||||||||||||||||||||||||||||||||
570 aLeuGlyAspAlaAlaThrProMetGluGlyProArgCysProProProA 587
```

```
426 ATGCAGCTCCTCTGCCCACTGTGTCTACTCTAAAGGACTTCAGAGAGGC  475
    |||||||||||||||||||||||||||||||||||||||||||||||||
735 spAlaAlaProLeuProThrValTyrSerLysGlyLeuGlnArgGly    751

476 TCTCCAGCAGGGCGCCTGGGACTCGGACCAAAATGGAACTCCAAGCGTGC 525
    ||||||||||||||||||||||||||||||||||||||||||||||||||
752 SerProAlaGlyAlaTrpAspSerArgGlnAsnGlyAsnGlyArgGlyAl 768

526 TTTGGGGACCCTGCCACTCCCACGGAAGGTCCTGCCGCCCACCTCCCC   575
    ||||||||||||||||||||||||||||||||||||||||||||||||
768 aLeuGlyAspProAlaThrProThrGluGlyProArgProProProA   785

576 GTCCTCCCTGTCGGCTCAGCATGGGCCGCCGTCACAACTCTGTAGCCCT  625
    |||||||||||||||||||||||||||||||||||||||||||||||||
785 rgProProCysArgLeuSerMetGlyArgArgHisLysLeuCysSerPro 801

626 GACCCGGGCCAGGCCAACAACAGTGAAGGCAGCAGCCATGACTACCTGCC 675
    ||||||||||||||||||||||||||||||||||||||||||||||||||
802 AspProGlyGlnAlaAsnAsnSerGluGlySerAspHisAspTyrLeuPr 818

676 CTTGGTGCGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACGCCGCCCT 725
    |||||||||||||||||||||||||||||||||||||||||||||||||||
818 oLeuValArgLeuGlnGluAlaProGlySerPheArgLeuAspAlaProP 835

726 TCTGCCCGCTGTGCCATCTCGCAGGAGGCCTCTGCCGTGCCTCCGCC    775
    |||||||||||||||||||||||||||||||||||||||||||||||||
835 heCysAlaAlaValArgIleSerGlnGluArgLeuCysArgAlaSerPro 851

776 TTTGCCCTGCCACCGCGCCAGGCCCTCAGCCCCACCTCGGCCTCATTGCCCTG 825
    ||||||||||||||||||||||||||||||||||||||||||||||||||
852 PheAlaValHisArgAlaSerLeuSerProThrSerAlaSerLeuProTr 868

826 GGCACTTCTGGGCCTTGGTGTTGGCCAGGTGACAGTGCCACGGCCTCCT  875
    ||||||||||||||||||||||||||||||||||||||||||||||||||
868 pAlaLeuLeuGlyProGlyValGlyInGlyAspSerAlaThrAlaSerC  885

876 GCCAGCGCCTCCCCCAGCTGCGGGCTCTCGAGGGGCCAGGTCAGGTGGACAGT 925
    |||||||||||||||||||||||||||||||||||||||||||||||||
885 ysSerProSerProSerGlySerGlyProGlyProGlyInValAspSer  901

926 GGGCGGGGCTCAGACACCGAGGCCTCCGATGGGCCGGAAGGGCTGGGCGG 975
    ||||||||||||||||||||||||||||||||::||||||||||||||||
902 GlyArgGlySerAspThrGluAlaSerGluGlyAlaGluGlyLeuGlyGl 918

976 CACCGACCTGCGGGGCCGGGGCCGACCTGGGCCACTGCCCTAGCACTCGCCTGGC 1025
    ||||||||||||||||||||||||||||||||||||||||||||||||||
918 yThrAspLeuArgGlyArgGlyArgThrTrpAlaThrAlaValAlaLeuAlaTrpL 935

1026 TGGACGACCGATGCGCGCCGCTGCCTTCGACGAGTGGGAACTGACACGCGGCC 1075
     ||||||||||||||||||||||||||||||||||||||||||||||||||
935 euGluHisArgCysAlaAlaAlaPheAspGluTrpGluLeuThrArgAlaAla 951

1076 AAGGCTGATTGCTGGCTGCGGGGCCCAGCACTGCCTGACGGCCTTGACCT 1125
     |||||||||||||||||||||||||||::|||||||||||||||||||||
952 LysAlaLeuAspCysTrpLeuArgGlyArgAlaGlnHisLeuProAspGlyLeuAspLe 968

1126 GGCCGCCCTCAAGGCCCAGGCCCGAGGGCTCTTCCTGCTACTGCGCCACT 1175
     |||||||||||||||||||||||||||||::::|||:::::|||
968 uAlaAlaLeuLysLysAlaAlaAlaArgGlyLeuPheLeuLeuLeuArgHisT 985

1176 GGGAACCAAAAACCTGCAGCTACACCTGTGTCTACAGCCCAGGCAACGTG 1225
     ||||||||||||||||||||||||||||||||||||||||||||||||||
985 rpAspGlnAsnLeuGlnLeuHisLeuCysTyrSerProAlaAsnVal    1001
```

```
seq_name: sp_mammal:Q9GMT9

seq_documentation_block:
ID   Q9GMT9      PRELIMINARY;      PRT;     803 AA.
AC   Q9GMT9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

seq_documentation_block:

```
ID   Q9GKV4                 PRELIMINARY;       PRT;    260 AA.
AC   Q9GKV4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 27.3 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051125; BAB18151.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   260 AA;  27336 MW;  02AEDD04E6EAF2ED CRC64;
```

alignment_scores:

```
          Quality:    1277.00
          Ratio:         5.170
Percent Similarity:     93.916      Percent Identity:   92.015
```

alignment_block:
US-09-786-136-4 x Q9GKV4

```
Align seq 1/1  to: Q9GKV4   from: 1  to: 260
                                       Length:     263
                                       Gaps:         1
```

seq_name: sp_mammal:Q9GKV4

```
                       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
231 ysAlaAlaAlaArgGlyLeuPheLeuLeuLeuArgHisTrpAspGlnAsn 247
                       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1187 CTGCAGCTACACCTGTCGTGCTACACCACCCCACGGAACGTG 1225

248 LeuGlnLeuHisLeuLeuCysTyrSerProAlaAsnVal 260
                       |||||||||||||||||||||||||||||||||||||||||||
1187                                                         1225

seq_name: sp_rodent:Q9CUE8

seq_documentation_block:
ID   Q9CUE8                    PRELIMINARY;       PRT;   1209 AA.
AC   Q9CUE8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4931403E03RIK PROTEIN (FRAGMENT).
GN   4931403E03RIK.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara T., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK016425; BAB30227.1; -.
DR   MGD; MGI:1922968; 4931403E03Rik.
DR   InterPro; IPR003015; HLH_Myc.
DR   InterPro; IPR002035; vWFA.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR   PROSITE; PS50234; VWFA; 1.
FT   NON_TER     1209     1209
SQ   SEQUENCE   1209 AA;  133186 MW;  87DFC69CC22O1142 CRC64;
```

alignment_block:
US-09-786-136-4 x Q9CUE8

Align seq 1/1   to: Q9CUE8   from: 1   to: 1209

Align seq 1/1   to: Q9CUE8   from: 1   to: 1209

alignment_scores:
```
         Quality:    359.50                 Length:        374
         Ratio:      1.892                   Gaps:          11
Percent Similarity:  50.802      Percent Identity:  30.749
```

```
239  GGCCATGCCCGGAGGTGCTGGCTTCAGCCCTTCAGACAACTAAGGTCAG 288
     |||||||                   ||||||||::::::::|||||||
874  GlySerAsnArgArgTyrGlnValAsnAlaValHisThrSerLysAlaCy 890

289  CTCTCGCCCCTCCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTA 338
     |||::.:                :::::::||||||||:::.::::
890  sSerValIleSerLysTyrTyrThrAlaPheValProValAspIleAsnLysA 907
```

```
339  GGGAGGTGCCTGCCTGGGGCCCTGCAGGTGTGCAGGTTGCCAGCTTCAGAGCCCGCTGAG 388
     |:.::: ||||||||||||    ||||||||||||||||:::||||
907  rgGlnTyrLeuProThrValValLysTyrProAsnSer................... 919

389  CCCCAGGAGAACCCCTCCTCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCT 438
     ||:::::||||||||||||||||||||:::::::||:||:::
919  erAlaGlyLeuGlyArgProGlnSerMetLeuArgGluHisSerArgAla 952

439  GCCCACTGTGTCTACTCTAAAAGGACTTCAGAGA............GGCTCTC 479
     ::::::::|||||||||::||||:::|||||||||           ||||||
920  .GlyAlaMetLeuSerPheArgAsnLeuThrArgGlnThrTrpGlyGlySers 936

480  CAGCAGGCGCGCTGG.................................GACTCGGACCAA 505
     ||||||||||||||                                  ||::::
936  erAlaGlyLeuGlyArgProGlnSerMetLeuArgGluHisSerArgAla 952

506  AATGGCAAACTCCAAG......................CGTGCTTTGGGAGCACCCTGCCCACTCC 546
     :::                   |||||||||     |||||||     |||
953  AlaGlyAspSerLysPheGlnThrLeuAlaLeuGlnAspSerProThrse 969

547  CACGGAAGGTCCTCGCCGCGCCCCACCTCCC............. 574
     |||                                         :::
969  rThrPheAsn.....LysThrProSerProGlyHisGluLysGlnThrThrA 985

575  .....CGTCCTCCCCTGTCCGGGCTCAGGATGGGCCGCGGCCCGTCACAAACTCTGT 619
     ||||||:|::|||||     :::|||||||||||     |||
985  laGluGlyProProGlnAsnLeuSerAlaSerAlaProSerSerMetLys 1001

620  AGCCCTGAC............................... 628
     ::: |                                   :::
1002 AlaThrGluThrLeuPheGlySerLysLeuAsnLeuAsnLysSerArgLe 1018

629  ...........................................CCGGGCCAGG 638
     ||||||||||||||||||||||||||||||||||||||||||    :
1018 uLeuThrArgAlaThrLysGlyPheLeuSerLeuProLysAlaS 1035

639  CCAACAACAGTGAAGGCAGC...GACCATGACTACCTGCCCCTTGGTGCGG 685
     ||||||||||||||||||||   ||||||||||||||||||||||||
1035 erGluAlaThrProGlySerGlnSerSerAspTyrIleProLeuValSer 1051

686  CTGCAGGAGGGACCAAGGCTCCTTCGCGCGTGACGGGCCCCTTCTCGCCGC 735
     ||||||||||||||||||||||||||||:||||||||||||||||||||
1035 ............................................ 1035

686  CTGCAGGAGGGACCAAGGCTCCTTCGCGCGTGACGGGCCCCTTCTCGCCGC 735
     |||||| ||| ||||:||||: |::||   ||||:||||||||||||
1051 LeuGlnLeuAlaSerGlyAlaPheLeuLeuAsnGluAlaPheCysThrh 1068

736  TGTGCCCATCTCGCAGGAGCCCTCTGCCGTGCCCTTTGCCGTGC 785
     :::::::||:|||||||||||||||:::||||||||||||||:::
1068 rIleGlnIleProMetGlyLysSerLysTrpThrSerProPheSerCysL 1085

786  ACCGGCGCCAGCCTCAGCCCCACCCTCGGCCTCATTGCCCTGGGCACTCTCG 835
     ::::: |||||||   |||||||||||:|||||||||||||||||||
1085 euArgMetSerLeuValThrArgGlnAspLeuLysThrGln......... 1099

836  GGCCCTGGTGTTGGCCAGGGTTGACAGTGACAACGGGCCCTTGCAGCCCGTC 885
     |||   :::: ::|||||||||||:::|||||||||||||||||||
1099 LeuGlnLeuAlaSerGlyAlaPheLeuLeuAsnGluAlaPheCysThrrh 1112

886  CCCCAACTCGGGCTCTGAGGGGCCAGGCCACCAGGTGGACAGTGGCCGGGGGCT 935
     :::::::::|||  |||||||||||||||:||||||||||||||||||||||
1112 osercysAspGlyIle......SerLeuLysSerGluGluSers 1125

936  CAGACACCGAGGCCTCCGATGGGGCCGGAAGGGCTGGGCGGCCACCGACCTG 985
     ||||| |||||||||||::||||||||||||||||||||||||||||||
1125 erAspGlnGluSerAsnAlaMetLeuGluHisMet.......... 1136

986  CGGGGCCGGGACCTGGGCCACTGCCTAGCACTCGCCTGGCTGGGAGCACCG 1035
     |||:::|||||||||:::|||||||||||||||||||||||||||||||
1137 ..GlyLysLeuTrpAlaThrValAlaIleAlaIleAlaTrpLeuGluHisSe 1152

1036 ATGCGCGCGTCCCTTCGACCGAGTGGGAACTGACACGGCCCAAGGGCTGATT 1085
     |||                                 |||||||||||||
1137 ..GlyLysLeuTrpAlaThrValAlaIleAlaIleAlaTrpLeuGluHisSe 1152

1036 ATGCGCGCGTCCCTTCGACCGAGTGGGAACTGACACGGCCCAAGGGCTGATT 1085
     |||  :::                 |||||||||||||||||||||||
1152 rSerAlaAsnTyrIleIleIleGluTrpGluValAlaAlaAlaLysAlaSers 1169
```

```

```
1086 GCTGGCTGCGGGCCCAGCACTTGCCTGACGGCCTTGACCTGGCCGCCCTC 1135
       |||:::::: ||| :::|||:::||| |||::::::|||
1169 erTrpValGluLysGlnLysValProGluGlyArgThrLeuSerThrLeu 1185

1136 AAGGCCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCACTGGGACCAAAA 1185
     ||| :::|||||| ||||||:::||||||||||||||||||:::::
1186 LysAsnThrAlaArgGlnLeuPheValLeuLeuArgHisTrpAspGluLy 1202

1186 CCTGCAGCTACACCTGCTGTGC 1207
     :|||::::::::::::||||||
1202 sLeuGluPheAsnMetLeuCys 1209

seq_name: sp_human:Q9BX42

seq_documentation_block:
ID   Q9BX42   PRELIMINARY;   PRT;   628 AA.
AC   Q9BX42;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   BA11M20.1 (TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR, RNA
DE   POLYMERASE II, C1, 130KD) (FRAGMENT).
GN   TAF2C1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Howden P.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137077; CAC36006.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER   628   628
SQ   SEQUENCE   628 AA; 59837 MW; ACFC31FCF698AC1C CRC64;

alignment_scores:
          Quality: 240.50   Length: 452
            Ratio: 1.215   Gaps: 26
 Percent Similarity: 43.805   Percent Identity: 27.434

alignment_block:
 US-09-786-136-4 x Q9BX42   ..

 Align seg 1/1 to: Q9BX42 from: 1 to: 628

 338 AGGGAGGTCCTGCCTGGGGCCCTG......CAGGTGTGCAGCTCAGAGCC 381
     ||| ::: ||||||||| :::||| |||::::::||
  51 ArgAlaAlaAlaAlaGlyAlaLeuGlyAsnHisValValSerGlySerPr 67

 382 CGCTGAGCCCCCAGGAACCCCTCCTGCCTCTCACAGCCATCTAGATGCAG 431
     |||| |||::: ||| ||||
  67 oAlaGlyAlaAlaGlyAlaGlyPro.....................AlaA 77

 432 CTCCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCA 481
     ||||| :::|||:::|||
  77 laPro.................................AlaGluGlyAlaPro 83

 482 GCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGG 531
     ::::::||| ::: |||:::::::::   ::: ||
  84 GlyAlaAlaProGluProProProAlaGlyArgAlaArgProGlyGlyGl 100

 532 GGACCCTGCCACTCCC...ACGGAAGGTCCTCGCCGCCCA.......... 568
     | ||| ||| :::||||||||||||
 100 yGlyProGlnArgProGlyProProSerProArgArgProLeuValProA 117

 569 ....CCTCCC............CGTCCTCCCTGTCGGCTCAGC 595
                |||||| ||||||||| ::: |||
 117 laGlyProAlaProProAlaAlaLysLeuArgProProProGluGlySer 133

 596 ATGGGCCGCCGTCACAAACTCTGTAGCCCTGACCCGGGCCAGGCCAACAA 645
        ||| |||:::||| |||::: |||
 134 AlaGlySer............CysAlaProValProAlaAlaAlaAlaVa 146

 646 CAGTGAAGGCAGCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGG 695
      ::: ||| :::
 146 lAlaAlaGlyProGlu.................................. 151

 696 CACCAGGCTCCTTCCGCCTGGACGCGCCCTTCTGCGCCGCTGTGCGCATC 745
       ||| ||||||::::::   ::::::|||::::::||| :::   ::
 152 ..Pro.AlaProAlaGlyProAlaLysProAlaGlyProAlaAlaLeuAl 167

 746 TCGCAGGAGCGCCTCTGCCGTGCCTCGCCCTTTGCCGTGCACCGCGCCAG 795
     : ||| |||::: ||| ||| ||| :::|||:
 167 aAlaArg...AlaGlyProGlyProGlyProGlyProGlyProGlyProG 183

 796 CCTCAGCCCCACCTCGGCCTCATTGCCCTGGG.................. 827
     ::   :::||| :::||| ||||||
 183 lyProGlyProGlyLysProAlaGlyProGlyAlaAlaGlnThrLeuAsn 199

 828 ........................CACTTCTGGGCCCTGGTGTTGGCCAG 853
                             ||| ||| :::
 200 GlySerAlaAlaLeuLeuAsnSerHisHisAlaAlaAlaProAlaValSe 216

 854 GGTGACAGTGCCACGGCCTCCTGCAGCCCGTCCC...CCAGCTCGGGCTC 900
      :::::: ||| ||| ||| ||| |||
 216 rLeuValAsnAsnGlyProAlaAlaLeuLeuProLeuProLysProAlaA 233

 901 TGAGGGGCCAGGCCAGGTGGACA.........GTGGGCGGGGCTCAGACA 941
          ||| ::: ||| ||||||:::|||
 233 laProGlyThrValIleGlnThrProProPheValGlyAlaAlaAlaPro 249

 942 CCGAGGCCTCCGATGGGGCGGAAGGGCTGGGCGGCACCGACCTGCGGGGC 991
     ||| ||| |||||||||::: ||
 250 ProAlaPro.....................AlaAlaProSerProProAl 259

 992 CGGACCTGGGCCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGC 1041
     |:::|||:::||| :::|||:::   ::: |
 259 aAlaProAlaProAla.........AlaProAlaAlaAlaProProProP 273

1042 CGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGGC 1091
     || |||:::
 273 roProProAla....................................... 276

1092 TGCGGGCCCAGCACTTGCCTGACGGCCTTGACCTGGCCGCCCTCAAGGCC 1141
           |||:::||| |||||| ||||||:::||| |||
 277 ......ProAlaThrLeuAla.ArgPro...ProGlyHisProAlaGly. 289

1142 GCAGCCCGAGGGCTCTTCCTGCTACTGCGCCA.................. 1173
          ||| ||| ||||||::: |||
 290 ..ProProThrAlaAlaProAlaValProProProAlaAlaAlaGlnAsn 305

1174 .........CTGGGACCAAAACCTGCAGCTACACCTGCTGTGCTACAGCC 1214
                 ||| |||||| :::|||||| :::   :::|
 306 GlyGlySerAlaGlyAlaAlaProAlaProAlaProAla.AlaGlyGlyP 322

1215 CAGCGAACGTGTGAAGGCTGCCC...CCTGCTGCTTGGGCTGGCGCCCCA 1261
     |||||:::||| ||| |||:::||| |||:::||||||
 322 roAlaGlyValSerGlyGlnProGlyProGlyAlaAlaAlaAlaAlaPro 338

1262 CCCAACACACTCAAGTCACTGCCGCCCAGGGCTGGCCTCTTGGTGCTGGG 1311
              :::|||::: |||:::
 339 AlaProGlyValLysAlaGluSerProLys.................... 348

1312 AAAGTGTAGGCTGGTGCCAGCCTGTCCCCCA................... 1342
            |||:::||| |||:::||||||
 349 ....ArgValValGlnAlaAlaProProAlaAlaGlnThrLeuAlaA 363

1343 .............................CTGCTTCTTACTCCCTCCCTG 1363
                                  ::::::::: |||::::::
 363 laSerGlyProAlaSerThrAlaAlaSerMetValIleGlyProThrMet 379

1364 GAGCCCTCTTGCCCCCACAAAAAGTGCCTGCCTGTGCTCTCTCCCTCTCC 1413
     :::   ::: ||| ::::::||| |||:::||
```

380 GlnGlyAlaLeuProSerProAlaAlaValV3lPro...... ProProAlaPr  394
        |::::
411 erLeuSerArgThrProThrAlaThrThrSerGlyIleArgAlaThrLeu  427
        |||::::||                    |||||||||||:::
1437 CCATCCTCTGAGCTCCCTGCAACACAGTGGAAGGTGAGAGGCCACAGTC 1486
        |||||                        ||:::
1437 oGlyThrProThrGlyLeuProLysGlyAlaAlaGlyIleAlaValThrGlnS 411
        ||||||
394 01-MAR-2001 (TrEMBlrel. 16, Created)
        |||||
1414 TCCCACCCCACTCACACTCCCC.................................T 1436
        |||||||
380 GlnGlyAlaLeuProSerProAlaAlaValValPro...... ProProAlaPr  394
```

seq_name: sp_plant:Q9FPQ6

```
seq_documentation_block:
ID   Q9FPQ6              PRELIMINARY;       PRT;     555 AA.
AC   Q9FPQ6;
DT   01-MAR-2001 (TrEMBlrel. 16, Created)
DT   01-MAR-2001 (TrEMBlrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBlrel. 19, Last annotation update)
DE   VEGETATIVE CELL WALL PROTEIN GP1.
GN   GP1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA   Goodenough U.W.;
RT   "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT   plant hydroxyproline-rich glycoproteins.";
RL   Biochemistry 0:0-0(2001).
EMBL; AF309494; AAG45420.1; -.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR029965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTNSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE  555 AA;  54219 MW;  6A584A90465502F5 CRC64;
```

alignment_block:
us-09-786-136-4 x Q9FPQ6

```
Align seg 1/1  to: Q9FPQ6  from: 1  to: 555

Quality:        235.00                      Length:      398
Ratio:          1.478                       Gaps:         20
Percent Similarity:  39.950     Percent Identity:  26.633
```

alignment_scores:

```
323 GTGGATGCTACTACTAGGGAGGAGGTCCTTGCCCTGGGGGCCCTGCAGGTGTGCAG 372
    ||||||||||:::                     |||||||||||||||||||
24  ValAlaSerAlaAsnAlaGlnCysValProGlyGlyIlePheAsnCysPr      40

373 CTCAGAGCCCCGCCTGAGCCCCCAGGAACCCCTCCTTGCCTCTCCACAGCCATC 422
    ::||||||||  |||       ||:||:||||||||
40  oProSerProAlaProSerProSerProAlaProPro................ 51

423 TAGATGCAGCTGCCTCTTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGA 472
    ::||:||||| |||                  |||||||
52  .....SerProAlaProPro......................... 56

473 GGCTCTTCCAGCAAGGCGGCCTGGGACTCGGACCCAAAATGGCAACTTCCAAGCG 522
    ||||||||||||                        |||||||
57  ...SerProAlaProPro................................ 61

523 TGCTTTGGGGAGACCCTGCCACTCCCACGGAAGGTCTTCGCGCGCCCA...C 569
    ::|||||||| |||||||                    |||
62  .........SerProAlaProSerProGlyProProSerProAlaP 75

570 CTCCCCGTCCTTCCCTGTGGCTCACGATGGCCGCGCGTCACAAACTCTGT 619
    |||||                        |||||||
75  roProSerProSerProAlaProPro................ 84
```

seq_name: sp_bacteria:Q9XDH2

```
620 AGCCCTGACCCGGGCCGCAGGCCAACAACAGTGAAGGCCAGCCACCATGACTA 669
    |||||||                    |||         :::
85  SerProAlaPro.ProSerProAlaProSerProAlaProProSerP 101

670 CCTGCCCTTGGTGCGGCGCTGCAGGAGGCACCAGGCTGCTTCCGCGCCTGGACG 719
    ::|||                        ||||||        :::
101 roAlaProProSerProAlaProProSerProAlaProProSerProAla 117

720 CGCCCTTCTCCGCCCCTGTGCGCCATCTCGCAGGACGGCCTCTGCCGTGCC 769
    |||||||||||||||||                    |||||||
118 ProProSerProPro..................SerProAlaProPr 127

770 TCGCCCCTTTGCCGTGCCACCGGCCACGCCCCACAGCCCTCAGCCCCCACCTCGGGCCTCATT 819
    |||    |||          |||::::    :::||||||||    |||
127 oSerProSerProProAlaProProSerProSerProSerProAlaP 144

820 GCCCTGGGCACTTCTGGCGCCCCTGGTGTTGGCCAGGGTGACAGTGCCACGG 869
    |||||                    |||||:::        :::
144 roPro...................LeuProSerProAla 151

870 CCTCCTGCAGCCCCGTGCCCCCAAGCCTCGGGGCTGAGGGGCCCAGGCCAGGTG 919
    ||||||:::                    ||||||||        :::
152 ProProSerProSerProProVal................ 159

920 GACAGTGGGCGGGGGGCTTCAGCACACCGAGGCCTTCCGATGGGGGGGGGGAAGGGCT 969
    |||||||                    |||||||||:::        :::
160 ..ProProSerProSerProProValProProSer. 170

970 GGGGCGCCACCGACCTCGCGGGGCGGGACCTGGGCACTGCCGTAGCACTCG 1019
    |||||||                    |||:::||    |||
171 .ProAlaProProSerProThrProProSerProSerProProValPro 186

1020 CCTGGCTGGAGCACCCCATGCCCGCCTGTGGGCCTTCGGACGAGTGGGAACTGACA 1069
    |||::::                    |||||||        :::
187 ProSerProAlaProProSerProAlaProProValProProSerProAl 203

1070 GCGGCCAAGGCTGATTGCTGGCTGCGGGGCCCAGCACTGCCTGACGGGCCT 1119
    |||                    |||||||
203 aProProSerProAlaProProValProProSerProAla......... 216

1120 TGACCTGGCGCCGCCCTCAAGGCGCCAGCCCGAGG.......GCTCTTCCTGC 1163
    |||||||                    |||||||:::    ||:::
217 ........ProProSerProPro.SerProAlaProProSerProProSe 230

1164 TACTGCGCCACTGGGACCAAAAAACCGCAGTTACACCTGCTGTGGTTACACGC 1213
    ||||||                    |||||:::        |||
230 rProAlaProProSerProSerProProAlaProProSerProValProP 247

1214 CCAGGCGAACGTGTGAAAGGCTTGCCCCCTGCTTCTTGGGCTGGGCGCGCCCACC 1263
    |||||||                    |||||||:::        |||
247 roSer.......ProAla.ProProSerPro.........AlaProPr 257

1264 CAACACACTCAAGTCAACTGCCGCCCACAGGGCTGGCCTTCTTGGTGCTGGGAA 1313
    |||:::                    |||||||
257 oSerProLysProProAlaProPro.............. 265

1314 AGTGTAGGCTGGTGCCAGCCTGTCCCCCACTGCTTTCTTACTCCCTG 1363
    ||||||                    |||||
266 ............ProPro..................ProSerPro 270

1364 GAGCCCTCTTGCCCCCCACAAAAAGTGCCTGGCTGTGCTCTCTCCCTCTCC 1413
    |||                    :::|||||||    ||
271 ProProProPro.ProAla.ProProPheProAlaAsnThrProMetPr 287

1414 TCCCACCCCACTCACACTCCCTCCAATCCTGTGAGCTCCCTGCCAACACAG 1463
    |||:::||||||                    |||||||
287 oProSerPro.Pro.....SerProProSerProSerProAlaProProThrPro 302

1464 TGGAAGGGTAGAGAGCCACAGTCCCCAAAATCCCTATG 1499
    :::||||                    ||:::||||
303 ProThrProProSerProSerProSerProProVal 314
```

seq_name: sp_bacteria:Q9XDH2

```
seq_documentation_block:
ID    Q9XDH2                    PRELIMINARY;      PRT;    763 AA.
AC    Q9XDH2;
DT    01-NOV-1999 (TrEMBLrel. 12, Created)
DT    01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT    01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE    PROLINE-RICH MUCIN HOMOLOG.
OS    Mycobacterium tuberculosis.
OC    Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC    Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX    NCBI_TaxID=1773;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=H37RV;
RX    MEDLINE=20090472; PubMed=10627046;
RA    Espitia C., Laclette J.P., Mondragon-Palomino M., Amador A.,
RA    Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT    "the PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RT    new family of fibronectin-binding proteins?";
RL    Microbiology 145:3487-3495(1999).
EMBL; AF071081; AAD41594.1; -.
DR    InterPro; IPR002951; Atrophin.
DR    InterPro; IPR003882; Pistil_extensin.
DR    InterPro; IPR002965; P_rich_extensn.
DR    PRINTS; PR01222; ATROPHIN.
DR    PRINTS; PR01217; PRICHEXTENSN.
DR    PRINTS; PR01218; PSTLEXTENSIN.
SQ    SEQUENCE   763 AA;  75035 MW;  39168EC45A5916F8 CRC64;
```

```
alignment_scores:
              Quality:    233.00
              Ratio:      1.131
Percent Similarity:      38.505          Percent Identity:  25.421

alignment_block:
US-09-786-136-4 x Q9XDH2

Align seg 1/1   to: Q9XDH2   from: 1 to: 763

                                                         Length:    535
                                                         Gaps:       27
```

```
214 CCACAGACTGCTCCTTGGTTC..........CACAGGCCATGCCCGAGGTGCT 257
    ||||||||||||||||||                |||  ||||||| ||||
23  ProLysSerLysProProPhePheProProAlaProProAlaProPro.CysT 39

258 GGCTTCGAGCCCTTCAGACAAGATAAGGTAGCTCAGCCTCGCCCCTCCTGCTTC 307
    |||||||||||  |||  ||| ||||||| ||||||||||||||
39  rpMet.........................LeuValSerAlaAlaProProCysPro 49

308 ACTTGC.................................CC 315
    ::|                                     ||
49  ProAlaProProAlaProProLysSerLysAlaProPheProPr 66

316 TGTAGCTGTTGGATGCTACTACTAGGGAGGTCCTGCCT...........  352
    |||||||||||    ::|||| ||||||
66  oValProProAlaAProProAlaArgGluLeuAlaProProLeuProPro 83

353 ..................GGGGCCTGCAGGTGTGC... 370
                       ||||||| ||||||||
83  laProProGluAlaProArgGluSerArgproAlaLeuProProCysPro 99

371 ............AGCTCAGAGAGCCCGCTGAGCCCCCAGG 396
              |||||||  |||||||
100 ProProProValIleProAspProProGluProAlaProPro.. 115

397 AACCCCTTCGTCTGCTCTCACAGC.............. 418
    ::|||||||||  ::|||
116 .ValProProAlaProAsnSerProProPhePheProProAlaP 132

419 ........CATCTAGATGCAGCTCTCCTGCCACTCGTTGTCTCACTCTAAAGGA 463
              |||           |||||||
132 roLysPheValProAlaProProValProProValProAlaProAsnSerProPro 148
```

```
464 CTTCAGAGAGGCGTCTCCACCAGCGGCGGCCCTGGGACTCGGGACCAAAATGGCCAA 513
    ::|||||||
149 PheProProPheProProAla................. 155

514 CTTCCAAGCCGTGCTTTGGGGGGACCCTGCCACTCCCACACGGAAGGTCCTCGCC 563
    ||||||| |||   |||||||  |||::|
156 .........AlaLeuAsnProProAlaProAlaProProLeuAlaa 169

564 GCCCACCTCCCCTCCCCTCCCTCGT.................... 586
    ::|||||||||||||||||||||||
169 snSerProProLeuProProAlaProProThrProAlaGlyThrProProPro 185

587 ................CGGCTCAGCAATGGGCGCCGGCCGTCACAAACTCTGTAG 621
                     ::|    ||| ||||||||  ::||
186 AlaAlaProTrpProProProValProAlaAlaProLysSerProProAlaSe 202

622 CCCTGACCCCGGGCCAGGCCAACAACAGTGAAGGACGACCCATGACTACC 671
    |||||||||||||||  |::|           |||
202 rProProArgProProAlaProProAlaThrProMetGluProPhe 219

672 TGCCCTTG...........................GTGCGGCTCCAGGAGGCACCA 700
    ||||||||                           |||
219 roProLeuProProValProProAspProIleSerLysGluThrProPro 235

701 GGCTTCCTTCCGCCTCGGACGGCCCCTTCTCGCCCGCCGCTGCGGCGCATCCCA 750
    |||||||     ||||||||        ||||||||||:::
236 .AlaProProAlaProProIleProProProAlaProVal.......... 247

751 GGAGGCGCTCTGCCGTGCCTCCGCCCCTTTCGCCGGTGCACCGCGCCCAGCCTCA 800
    |||||||||||  |||::|||||||||||  ::|||||||||        ::
248 .ProIleProProValPro...ProLeuPro....ProValProAsnLys 261

801 GCCCCACCTCGGCGCTCTCATTGCCCTGGGCACTTCTGGGCCCCTGCGTGTTGGC 850
    ||||||||||                        |||:::::||||||::||
262 IleProProAlaProProAlaProProValAlaValAlaLeuAlaAlaValLeuVa 278

851 CAGGGTGACAGTGCCACGGGCCTCCTGCCAGCCCCGT.................C 885
    ||||           ||||||||||||||||||                    |
278 lAlaProCysProProLeuProProLeuProAsnAsnHisProProAlaP 295

886 CCCCAGCTCGGGCTCTGAGGGGCCAGGCCAGTGGACAGTGGGCGGGGCT 935
    |||||||||      ||| ||||||||||||||||||||||||||
295 roProAlaAlaProValProGlyValProLeuValAlaProLeuProAsnSer 311

936 CAGACACCGAGGGCCTCCGATGGGGCGGGAAGGGCTTGGGGGCACCGACCTG 985
    |||||||||||||     |||||||||||||||||||||||
312 HisProProAlaProPro................SerAlaProValPr 322

986 CGGGGCCGGACCTGGGCCACTGCCGTAGCAC.................. 1016
    ::|:::      ::||||||||||||||
322 oGlyValProLeuValAlaProLeuProIleSerGlyAlaArgProValSerValr 339

1016 .................TCGCCTGGCTGGAGCACCGATGCGCC 1042
                     ::|||||  ||||||| |||
339 rpLysProGlySerPheThrThrLeuSerThrPheCysCysArgValCysSer 355

1043 GCTGCCTTCGACGAGTGGGAAACTGACAGCGCCAAAGGTGATTGCTGGCT 1092
    |||||||  |||||||||||||||||||||||||||||||||
356 GlyGluValLeuValAlaGlyValAlaLeuAsnProSerArgProSerArgSerPr 372

1093 GCGGGCCCAGCACTTGCCTGACGGCCTG..............ACTTGGCCG 1130
    |||                                        |||
372 oLeuThrThrThrThr.............ProAlaLeuProAlaProI 384

1131 CCCTCAAGGCCCGCAGCGCCCGAGGGCT............CTTCCTG 1162
    ||||                                     ||||||
384 leProProLeuProProLeuProProLeuProIleGlnAsnThrAlaValPro 400

1163 CTACTGCCGCCACTGGACACCAAAAACCTGCAGCTACACCTGCT..GTGCTA 1209
    |||                                           ::
401 ProIleProProLeuProPro..ValThrAlaLeuAlaProProLeuProP 417
```

```
seq_name: sp_plant:Q9SPM0

seq_documentation_block:
ID    Q9SPM0;                    PRELIMINARY;      PRT;    1315 AA.
AC    Q9SPM0;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE    EXTENSIN-LIKE PROTEIN.
GN    PEX2.
OS    Zea mays (Maize).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC    Panicoideae; Andropogoneae; Zea.
OX    NCBI_TaxID=4577;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=POLLEN;
RX    MEDLINE=21330248; PubMed=11437249;
RA    Stratford S., Barne W., Hohorst D.L., Sagert J.G., Cotter R.,
RA    Golubiewski A., Showalter A.M., McCormick S., Bedinger P.;
RT    "A leucine-rich repeat region is conserved in pollen extensin-like
RT    (pex) proteins in monocots and dicots.";
RL    Plant Mol. Biol. 46:43-56(2001).
DR    EMBL; AF159297; AAD55980.1; -.
DR    InterPro; IPR001611; LRR.
DR    InterPro; IPR003592; LRR_out.
DR    InterPro; IPR002965; P_rich_extensn.
DR    InterPro; IPR001998; Xylose_isom.
DR    Pfam; PF00560; LRR; 3.
DR    PRINTS; PR01217; PRICHEXTENSN.
DR    SMART; SM00370; LRR; 3.
DR    PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
SQ    SEQUENCE  1315 AA;  134402 MW;  64C97A2A01F0936F CRC64;
```

```
alignment_scores:
                  Quality:      226.50          Length:      510
                  Ratio:          1.049         Gaps:         29
Percent Similarity:  42.353     Percent Identity:  25.490
```

```
alignment_block:
US-09-786-136-4:
Align seq 1/1   to: Q9SPM0
US-09-786-136-4 x Q9SPM0

    211 GCTCCACAGACTGCTTCCTTGGTTCCACAGGCCATGCCCGAGGTGCTGGC    260
        ||||||       ::||||| ||||   || :::::::||
    812 AlaProValSerSerPro............................ProGlnValGlyLy   822
```

```
1078 ArgThrArgValSerSerProArgProArgPro...ValValLysCysProPr 1093
     |||            :::                    :::::         |||
1100 CAGCACTTGGCGTGACAGGGCCTTGACCTGCCGCCCCTCCAAGGCCGCCAGCCCG 1149
     |  |||          |:::::||           ||||||| |||  :::|||
1093 oProThrLeuValSerSer......Pro......Pro.AlaPro 1104

1150 AGGGCTCTTCCTCCTCCTACTGCGCCACTG 1176
     ||||||
1105 LysSerLeuProProProThrProValSerSerProSerProProGluValy 1121

1177 ...............................GGACCAAAA 1185
     :::|||||
1121 sSerSerProProProThrProValSerSerProSerProProAlaProLys 1138

1186 ....CCTGCAGCTACACCTGCTGTCTACAGCCCAGCGAACGTGTGAAG 1230
     ||||::||| |||||||:::
1138 erSerProProProThrProValSerSerProProGluLeuLysSer 1154

1231 GCTGCCCCCTGCTGCTTGGGCGTCGGCGCCCCACCCAACACACTCAAGTCAC 1280
     ::: ||||||        ||||::::||::::::||::
1155 Ser.ProProProAlaProValSerSerProSerProSerAlaProLysSerS 1171

1281 TGCGGCGCCCAGGGGCTGGCCTC...TTGCTGCTGGGAAAGTGTTAGGCTGGTG 1327
     |||||||      :::          |||   ||
1171 erProProProAlaProValSerLeuProProProGluValValLysSer 1187

1328 CCAGGCCGTCCCCACCACTGCTTCTTACTCCC...............TCCGGA 1365
     |||||||::                                         |||
1188 ProProProAlaProIleSerSerSerProProProAlaProVal... 1204

1366 GCCCTCTTGCCCCCACAAAAAGTGC......CTGCCTGTGCTGCTCTCCCCTCTC 1412
     |:::                      ||||||      ||||||
1204 oProProAlaProMetSerSerLeuProProProValLysSerProProP 1221

1413 CTCCCACCCCCACTCACACTCCCCTCCA..................... 1439
     |||::|||:::|||||:::
1221 roProProAlaProValSer.SerProProProMetLysSerProProPr 1237

1440 ..............TCCTCTGGAGCTCCCCTGCAACACAGTGGAAGGGTAGAG 1476
     |||||||::                         |||||||
1237 oProProAlaProIleSerSerSerProProProAlaProVal....... 1251

1477 AGCCACAGTCCCCAAAATCCT 1496
     |||    |||   |||
1251 roProSerLeuProProPro 1257
```

alignment_scores:
```
          Quality:  224.00          Length:   498
          Ratio:      1.093         Gaps:      32
Percent Similarity:   41.165    Percent Identity:   27.309
```

alignment_block:
US-09-786-136-4 x O41935

```
Align seg 1/1  to: O41935  from: 1  to: 585
               US-09-786-136-4  from: 1  to: 585

238 AGGCCATGCCCCGGAGGGTGCTGGCTTCGAGCCCTTCAGACAAGTAAGGTCA 287
    ||||||  |||||  |||||||::::
141 ArgProProProGluLeuGlyProGlySerProThrSerPro......... 154

288 GCTCTGCCCCCTCCTGCTTCACTTGCCCTGTAGCTGTGGATGGTTACTPACT 337
    |||        :::|| |||  ||
155 ............AlaProSerArgAlaGly.AlaArgIle 163

338 AGGGAGGTGCCTGCTGGGGCCCGTGCCAGGTGTGCAGCTCAGAGCCCGCTGA 387
    :::  ||||    |||     |||  ||  :::
164 ProAsp...LeuProGlyProLeuProSerTrpGlyProAspPro...Ar 178

388 GCCCCCAGGAACCCCTCCTTGCCTCTCACAGCCATCTAGATGCAGCTCCTC 437
    |||::::|| |||||||||||   ||||||||||  |::::||
178 gProProArgProProProGluLeuGlyProGlyLysSerProThrSerProA 195

438 TGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGC 487
    |||         ||||  |||||||||  ||||||||  :
195 laPro.............SerArgAlaGly 200

488 GCCTGGGACTCGGGACCAAAAATGGCAACTCCAAGCGTGCTTTGGGGGACC 537
    |||          |||     ||||||||||||||| |:::
201 AlaArgIleProAspProLeuProGly.................. 208

538 TGCCACTCCCACGGAAGGTCCTCGCGCCCCACCTCCCGGTCTCCCCGTGTC 587
    ||||||||||||||||||   |||||||||||||:::
209 .ProLeuProSerTrpGlyProAspProArgProProArgProProProG 225

588 GGCTCAGCATCGGGCGCCGGGCCGTCACAAACTGTAGCCGTGGACCCGGCCAG 637
    :::|||||| ||||||||||||||||  |||::::
225 luLeuGlyProGly.........SerProThrSerProAlaProSerArg 238

638 GCCAACACAGTGAAGGCAGCGACCATGACTACCTGCCCCTTGGTGCGGCT 687
    |||::::                         ::||||||||
239 AlaGly............................AlaArgIl 243

688 GCAGGAGGCCACCAGGCTCCT...TCCGCCTGG 716
    |||:::|                 ::||||
243 eProAspLeuProGly.ProLeuProSerTrpGlyProAspProArgPro 259

717 ..........................ACGCGCCGTGCGGCCGATCTCG 748
    |||:::::|||     :::|||
260 ProArgProProProGluLeuGlyProGlySerProThrSerProAlaPr 276

749 CAGGAGCGCGCTCTCGCC...............GTGCCTCGCCCTTTGCCGTGCA 786
    |||::::|||           :::|||||||||||
276 oSerArgAlaGlyAlaArgIleProAspLeuProGlyProLeuProSerT 293

787 CCCGCGCCAGCCTCAGCCCCACCTCGGGCCTCATTGCCCTCGGGCACTCTGG 836
    |||||||||||||| |||||||||||||||
293 rpGlyProAspProArgProProArgPro......ProProGlu..LeuG 307

837 GCCCTGGTGTGTTGGGCCCAGGGTGACAGTGCCACGGCCGCCTCCCAGCCCGTCC 886
    ||||||||       |||::: |||||||||||||||| |:::
307 lyProGly..........SerProThrSerProAla 315
```

```
seq_name: sp_fungi:O42854

seq_documentation_block:
ID   O42854      PRELIMINARY;      PRT;   1611 AA.
AC   O42854;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 170.5 KDA PROTEIN.
```

```
GN   SPAC23A1.17.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H';
RA   Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021813; CAA16991.1; -.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1611 AA;  170526 MW;  B3DEF9F83C1A7542 CRC64;
```

```
alignment_scores:
                     Quality:  218.00          Length:    460
                     Ratio:    0.991           Gaps:       26
Percent Similarity:  47.826    Percent Identity:  26.957

alignment_block:
US-09-786-136-4 x O42854

Align seg 1/1   to: O42854   from: 1  to: 1611
```

```
 887 CCCACCTCGGGCTCT...............GAGGGGGCCAGGGCCAGGTGGGACAG 924
     ||||||                         ::|    |||||    ||  :::||
 316 ProSerArgAlaGlyIleProAspLeuProGlyProLeuProSe            332

 925 TGGGCGGGGGGCTCAGACACCGAGGGCCTCCGATG..........GGGCGGGAAGGGC 968
     ||||||                                        :::  |||||
 332 rTrp...GlyProAsp.ProArgProProArgProProGluLeuGly            347

 969 TGGGGCGGCACCGACCTGCGGGGCCGGA 995
     :::::  |||||||||||       |||
 348 ProGlySerProThrSerProAlaProSerArgAlaGlyAlaArgIlePr 364

 996 CCTGGGCCACTGCCGTACCACTGCCGTCTGCCTGGAGCACCGAT. 1037
     |||||||||||||||                    ||||||
 364 oAspLeuProGlyProLeu..............ProSerTrpGlyProAspP 377

1037 ............... 1037
     ...............
 377 roArgProProArgProProGluLeuGlyProGlySerProThrSer 393

1038 .........................GCGCCGCT 1045
     ::|||||||
 394 ProAlaProSerArgAlaGlyAlaGlyIleAlaArgIleProAspLeuProGlyProLe 410

1046 GCCTTCGACCAGTGGGAACTGACACCGGCCAAGGCTGATTGCCTGCTGCTGCG 1095
     |||||||                 :::     ||||||
 410 uProSerTrpGlyProAspProArgProArgProProGluLeuG 427

1096 GGGCCAGCACTTGCCTGACGGGCCTTGACCTGGCCGCGGCCCTCAAGGCCGCAG 1145
     |||||||
 427 lyProGlySerProThrSer.ProAlaProSerArgAlaGlyAlaGlyAlaArgIl 443

1146 .........GCTCTTCCTGCTCTACTGCGCCACTGG 1177
     |||                    |||
 443 eProAspProSerArgProGlnLeuGlyProGlySerProArgP 460

1178 GACCAAAACCTGCACTACACCTGCTGTGTCTACACCCC...........AGCGAACGT 1224
     |||                                              :::
1178                          |||
 460 heProArgProSerArgProGlyProProGlnLeuGlyProGlySerProAsp 476

1225 GTGAAGGCGTGCCCCCTGCGTGCTGCTTGGGGCTGGCCACCCAACACACTCA 1274
     |||                          |||              |||
 477 LeuProGly.ProLeuProSerTrpGlyProAspProProThrPhe.... 491

1275 AGTCACTGCGCCCCCAGGGCGTGGCCCTCTTGGTGCTGGAAAGTGTAGGCCTG 1324
     |||                     |||
 492 .ProAlaProSerArgAlaGly..............AlaArg 500

1325 GTGCCAGCCGTGTCCCCACTGGCTTCTTACCCCTCCCTGAGCCCTCTTG 1374
     :::|||   |||||||||:::                     |||
 501 IlePro.ProValProProThrPhe..............ProA 510

1375 CCCCCACAAAAGTGCCCTCCGCTGTGCTCTCTCCCTCTTCCCACCCCAC 1424
     |||
 510 laProSerArgAlaGlyAlaArgIleProArgProProArgProPro 526

1425 TCACACTCCCCTCCATCCTCTGAGCTCCCTGCA.......ACACAGTGGAA 1468
     :::  |||              ||||||||||       ||| :|||
 527 GluLeuGlyProGlySerProAspLeuProGlyProLeuProSerTrpG1 543

1469 GGGTAGAGAGCCACAGTCC....CCAAATCCT 1496
     |||         |||||:::    |||:||||
 543 yProValProProProSerArgProSerPro 553
```

```
 314 CCTGTAGCT............GTTGGATGCTACTACTAGGGAGGAGGTTCCTGCCCTGGGGCCCT 360
     ||||||||                :::|||  ||||::: || ||
 812 ProValSerIleValThrSerGlyLysGlyArgProAlaLeuProAspGluMe 828

 361 GCAGGTGTGCCAGCTCAGAACCCCCTGACCCC....CCAGGAACCCCCTCCTTG 407
     ||||||:::                           ||:|| :::
 828 tAlaSerProSerSerIleGlyHisProLeuProSerProProProProA 845

 408 CCTCT............................CACAGGCCACTA 424
     ||:::                            |||||:|||
 845 laAspPheAsnSerLeuAsnValAspPheTyrGluProHisSerTyrLeu 861

 425 GATGCA...GCTCCTCTGCCC..................... 442
     ||       |||||||
 862 GluSerProAlaProGlnProSerTyrGluGluSerPheAs 878

 443 ......ACGTTTGTCTACTTCTAAAAGGACTTCAGAGAGGGCTCTCAGCAGGCG 488
     |||||||:::          :::          ||| :::
 878 nAlaThrValIleHisAlaProThrProSer.ThrAlaThrPheGlnGly 894

 489 CCTGGGACTCGGACCAAAATGGCAAACTCCAAGGGTGCTTTGGGGGACCCT 538
     ||||||                      :::||||||||||
 895 HisProThrIleSerGluAsnValAlaThrPro 904

 539 GCCCACTCCCACGGAAGGTGTTCGCGCCCGCCACCTCCCCGTCCCCTGTCG 588
     |||         |||||::::                ||| :::|||||
 905 .ProLeuLysGlnAspValThrGluSer......LysAlaSerProValA 919

 589 GCTCAGCATGGGCGCCGGTCACAAAACTCTGTTAGCCCTGACCCGG...... 632
     :::::|||||         |||:::|||||||||:::||||||
 919 laAspAlaSerAlaThrHisGlnSerSerThrGlyLeuThrGlnGluIle 935

 633 ...GCCAGGCCAACAACAGTGAAGGCAGCGACCCATGACTACTTGCCCTTG 679
     ::::::::::||||                  ||| |||||
 936 ThrGlnLeuGlySerAsnMetArgLeuProThrLysLeuThrArgProSe 952

 680 GTGCGGGCTGCCAGGAGGCACCAGGCTCCTTCCGCGTGGACGCGCCCTTCTG 729
     ||||||:::           |||||||:::||
 952 rAsnAspGlyArgLysAlaSerGlyPro 965

 730 CGCCGCGTGTGCCCATCTCGCAGGAGCCCCTCGCCGTGGGCCCGCCCTTTG 779
     |||||                    |||||||
 965 ArgProAlaA 965
```

```
 965 laPro..............................................ProSerIleProProLeu  973

 780 CCG..............TGGACCGGGCCCAGCCTCAGCCCCACCTCGGCCTCA  817
     |||                :::::||||||||:::||||||||:::
 974 ProValSerAsnIleLeuSerSerProThrSerGluProProLysAspHi  990

 818 TTGCCCTGGGCACTTCTGGGCCCCTGGTGTTGGGCCAGGGTGACAGTGCCAC  867
     |     |||||:                |||       |||  |||
 990 sProProSerAla.....ProLeuSerLysProValSerThrSerProA 1005

 868 GGCCTTCCTGCAGCCCGT...CCCCAGCGTCGGGGGCTTCTGAGGGGCCAGGCC  914
     |||||  |||||:::|   ||||||||:::     :::|||||||||
1005 1aAlaProLeuAlaArgValProProValProLysLeuSerSerLysAla 1021

 915 AGGTGGACAGTGGGCGGGGGCTCAGACACCGAGGCCTCCGATGGGGCCGGAA  964
     ||||||||:::
1022 ...................ProProVal.......... 1024

 965 GGGCTGGGGGCACCGACCTGCGGGGCCCGGACCTGGGCCACTGCCGTAGC 1014
     ||||:|::
1025 .......ProLeuProSerAlaAspAlaProProIleProValPro...... 1037

1015 ACTCGCCTGGCTGGAGCACCGATGCGCCCGCTGCCTTCGACGAGTGGGAAC 1064
     ||||||||:::   ||:::|||
1038 ...............SerThrAlaProProValProIleProThrSerThr 1049

1065 TGACAGCGGCCCAAGGCTGATTGCTGGGCTGC.................. 1094
     |||:::
1050 ProProValProLysSerSerSerGlyAlaProSerAlaProProVa 1066

1095 ...GGGCCCAGCACTTGCCTGACGGGCCCTTGACCTGGCCGCCCCTCAAGGC 1140
     ::|||||||||:::      :::  |||:::|||
1066 1ProAlaProSerSerGlyIleIleProSerIle.ProAla..ProSerGly 1081

1141 CGCACCCGAGGGGCTCTTCCTGCTACTGGGCCCACTGGGACCAAAAACCTGC 1190
     |||
1082 AlaProProValProAlaProSerGlyIleIleProProValProLysProSe 1098

1191 A...GCTACACCTGCTGGTGCTACAGCCCAGCGGAACGTGTGAAGGCTGCCC 1237
     :||:::|||
1098 rValAlaAlaProProValProLysProSerValAlaValProProValP 1115

1238 CCTGCTGCTTGGGGTGGGCGCCCCACCCAACACACTCAAGTCACTGCC... 1284
     |||:::|||    :::||||:::|||
1115 roAlaProSerGlyIleAlaProProValProLysProSerValAlaAlaPro 1131

1285 ..............GCCCAGGGCTGGCCT.................CTTCGG 1304
     ::|||   |||
1132 ProValProValProSerGlyIleAlaProProValProLysProSerValAl 1148

1305 TGCTCGGAAAGTGTRAGGCTGGTGCCAGCCTGTCCCCACTGCTTCTTTACT 1354
     ||||||
1148 aAlaProProValProAlaProSerGlyIleAlaProProValProLysProS 1165

1355 CCCTC.............CCTGGAGCCCTCTTGCCCCCACAA 1383
     |||||            :::|||||||
1165 erValAlaAlaProProValProAlaProSerSerGlyIleProProVal 1181

1384 AAAGTGCCTGCCCTGTGCTCTCCTCCTCTCCCACCCCACTCACACTCC 1433
     |||||||:::     :::||||:::
1182 ProLysProAla.AlaGlyValProProValProProSerGluAla 1197

1434 CCT................CCATCCTCTGAGCTCCCTGCAACACAGTGGAAGGG 1471
     :|||               :::|||
1198 ProProValProLysProSerValGlyValProProProValProProProSe 1214

1472 TAGAGAGCCACAGTCCCCAAAATCCT 1496
     |||   |||:::||||/
1214 rThrAlaProProValProThrPro 1222
```

```
OM of: US-09-786-136-4 to: A_Geneseq_032802.*    out_format : pfs

Date: Sep 19, 2002   5:41 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09786136/runat_16092002_141242_19172/app_query.fasta_1.1592
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-NS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-MAX=100 -THR_MIN=0 -ALIGN=10 -MINLEN=0 -MAXLEN=2000000000
-M=ext -HEAPSIZE=500 -MINLEN=0 -NCPU=6 -ICPU=3 -LONGLOG
-USER=US09786136_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-136-4
Query length: 1611
Database: A_Geneseq_032802.*
Database sequences: 747574
Database length: 111073796
Search time (sec): 85.420000
```

score_list:							
Sequence	Strd	Orig	ZScore	EScore	Len	Documentation	
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82530	+	1122.00	1466.77	7.0e-74	210		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW31852	+	233.00	290.67	6.3e-09	76		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW31855	+	227.50	286.09	1.6e-08	57		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB11413	+	227.00	287.80	1.6e-08	44		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW72204	+	222.00	263.45	8.0e-07	31		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03533	+	220.50	272.79	5.2e-08	90		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY50558	+	220.00	279.09	5.0e-08	41		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03533	+	218.50	270.17	7.3e-08	90		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	+	217.00	265.50	9.9e-08	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504	+	216.50	266.06	1.1e-07	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70063	+	215.00	271.98	1.2e-07	44		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB71434	+	215.00	257.12	1.5e-07	22		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21932	+	212.00	267.57	2.0e-07	47		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM52322	+	210.50	263.79	2.6e-07	57		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAC67370	+	210.50	263.79	2.6e-07	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26240	+	210.50	258.03	2.9e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239	+	210.50	256.84	3.0e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08112	+	209.50	257.31	3.4e-07	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB11527	+	208.00	255.39	4.4e-07	10		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	+	207.50	254.27	4.8e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	+	207.50	253.06	4.9e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB11013	+	207.00	258.71	4.7e-07	60		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR55248	+	205.50	249.64	7.3e-07	13		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB11013	+	205.00	256.09	6.6e-07	60		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY28843	-	205.00	255.58	6.7e-07	64		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95856	-	205.00	255.58	6.7e-07	64		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB62332	-	205.00	255.58	6.7e-07	64		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:ABG03530	-	204.50	255.54	7.2e-07	59		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26240	+	204.50	249.52	8.6e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239	+	204.00	248.33	8.8e-07	12		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14000	+	201.00	250.97	1.3e-06	59		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:ABG03530	+	200.00	249.65	1.5e-06	59		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14738	+	200.00	247.23	1.6e-06	78		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03868	+	198.50	245.52	2.1e-06	76		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03569	+	197.50	242.87	2.7e-06	81		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM24516	+	196.50	225.44	4.0e-06	51		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAG38942	+	195.00	240.90	3.7e-06	76		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95541	+	194.50	236.69	4.3e-06	11		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAA41231	-	194.50	227.88	5.0e-06	29		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23029	-	194.50	226.03	5.2e-06	36		
```

```
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB50363  +  193.50  226.56  6.0e-06
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB50363  +  193.50  226.13  6.0e-06
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14126  -  192.50  238.62  5.5e-06
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08321  -  191.50  239.66  6.3e-06
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW81589  +  191.00  241.83  6.5e-06
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82530

seq_documentation_block:
ID   AAY82530 standard; Protein; 210 AA.
XX
AC   AAY82530;
XX
DT   13-JUL-2000  (first entry)
XX
DE   Human neurotransmitter associated protein sequence SEQ ID NO:6.
XX
KW   Human; neurotransmitter; diagnosis; antiparkinsonian; neuroleptic;
KW   anticonvulsant; nootropic; tranquilizer; neuroprotective; cytostatic;
KW   antidepressant; antidiabetic; gynaecological; immunosuppressive;
KW   neurological disorder; hormone-related disease; Parkinson's disease;
KW   schizophrenia; epilepsy; female reproductive disorder;
KW   attention deficit disorder.
XX
OS   Homo sapiens.
XX
PN   WO20012685-A2.
XX
PD   09-MAR-2000.
XX
PF   26-AUG-1999;   99WO-US19615.
PR   01-SEP-1998;   98US-0144952.
PR   01-SEP-1998;   98US-0155194.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Walker MG, Volkmuth W, Klingler TM;
XX
DR   WPI; 2000-375619/32.
DR   N-PSDB; AAA08405.
XX
PT   A new purified polynucleotide comprising a gene that is coexpressed
PT   with neurotransmitter-processing-specific genes in biological samples
PT   for diagnosing, treating or preventing neurological and hormone-related
PT   diseases -
XX
PS   Claim 4; Page 35-36; 36pp; English.
XX
CC   The present invention describes a purified polynucleotide comprising a
CC   gene that is coexpressed with one or more neurotransmitter-processing-
CC   specific genes in biological samples. The neurotransmitter-processing-
CC   specific genes are L-tyrosine hydroxylase (TH), aromatic amino acid
CC   decarboxylase (AADC), dopamine beta-hydroxylase (DBH), nicotinic
CC   acetylcholine receptor alpha3 subunit precursor (nAchR-alpha3),
CC   secretogranin I and II, Rab3a, human cocaine and amphetamine regulated
CC   transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), and ARIX
CC   homeodomain protein. The present sequence represents a neurotransmitter
CC   associated protein sequence from the present invention. The
CC   polynucleotides from the present invention can have antiparkinsonian,
CC   neuroleptic, anticonvulsant, nootropic, tranquillizer, neuroprotective,
CC   cytostatic, antidepressant, antidiabetic, gynaecological and
CC   immunosuppressive activities. They can be used for diagnosing, treating,
CC   preventing or evaluating therapies for neurological and hormone-related
CC   diseases, particularly Parkinson's disease, schizophrenia, epilepsy,
CC   female reproductive disorders and attention deficit disorder. The gene
CC   products are therapeutic proteins and targets of therapeutics against
CC   the diseases.
XX
SQ   Sequence   210 AA;
XX
alignment_scores:
```

```
       Quality: 1122.00          Length:     210
         Ratio:    5.343           Gaps:       0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-786-136-4 x AAY82530    ..

Align seg 1/1  to: AAY82530   from:  1  to: 210

 596 ATGGGCCGCGTCACAAACTCTGTAGCCCTGACCCGGCCAGGCCAACAA 645
     ||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetGlyArgHisLysLysLeuCysSerProAspProGlyGlnAlaAsnAs 17

 646 CAGTGAAGGCAGGCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGG 695
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  17 nSerGluGlySerAspHisAspTyrLeuProLeuValArgLeuGlnGluA 34

 696 CACCAGGCTCCTTCCGCCTGGAGCGCCCCTTCTGCCGCGTGCCGCATC 745
     ||||||||||||||||||||||||||||||||||||||||||||||||
  34 laProGlySerPheArgLeuAspArgProPheCysAlaAlaAlaValArgIle 50

 746 TCGCAGGAGGCGCCCTGCCCTCGCCCTTTGCCGTGCACCGCCGCCAG 795
     |||||||||||||||||||||||||||||||||||||||||||||||
  51 SerGlnGluArgLeuCysArgAlaSerProPheAlaValHisArgArgAlaSer 67

 796 CCTCAGCCCCCACCTGGCCCTCATTGCCCTGGCACTTCTGGGCCCTGGTG 845
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 rLeuSerProThrSerAlaSerLeuProTrpAlaLeuLeuGlyProGlyV 84

 846 TTGGCCAGGTGACAGTGCCACGGCCGTCCTCCAGCCGCCCCCAGCTCG 895
     ||||||||||||||||||||||||||||||||||||||||||||||||
  84 alGlyGlnGlyAspSerAlaThrAlaSerCysSerProSerProSerSer 100

 896 GGCTTCGAGGGCCAGGCCAGTGACAGTGGCCGGGCTCAGACACCGA 945
     ||||||||||||||||||||||||||||||||||||||||||||||
 101 GlySerGluGlyProGlyGlnValAspSerGlyArgAlaGlnThrProThrGl 117

 946 GGCCTCCGATGGGGCGGAAGGGCTGGGCGGCCACCGACCTGCGGGGCCGGA 995
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 117 uAlaSerAspGlyAlaGluGlyLeuGlyGlyThrAspLeuArgGlyArgT 134

 996 CCTGGGCCCACTGCCCTAGCACTCGCCTGGCTGGAGCACCGATGCGCCGCT 1045
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 hrTrpAlaThrAlaValAlaLeuAlaTrpLeuGluHisArgCysAlaAla 150

1046 GCCTTCGACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGCGTGCC 1095
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 AlaPheAspGluTrpGluLeuThrAlaAlaLysAlaAlaAspCysTrpLeuGluAr 167

1096 GGCCCAGCACTTGCCTGACGGCCGTTGACCTGGCCGCCCTCAAGGCCCAG 1145
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 gAlaGlnHisLeuProAspGlyLeuAspLeuAlaAlaLeuLysAlaAlaA 184

     CCCGAGGGGCCTCTTCCTGCTACTGCGCCACTGGGACCAAAACCTGCAGCTA 1195
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 184 laArgGlyLeuPheLeuLeuLeuArgHisTrpAspHisTrpAspGlnAsnLeuGlnLeu 200

1196 CACCTGCTGTTGCTACAGCCCAGCGAACGTG 1225
     |||||||||||||||||||||||||||||||
 201 HisLeuLeuCysTyrSerProAlaAsnVal 210
```

seq_name: /SIDS1/gcgdata/geneseqp/geneseqp-embl/AA1997.DAT:AAW31852

seq_documentation_block:
```
ID   AAW31852 standard; Protein; 763 AA.
XX
AC   AAW31852;
XX
DT   27-APR-1998   (first entry)
XX
DE   Mycobacterium tuberculosis 74 kDa protein.
XX
KW   Tuberculosis; mycobacteria; infection; diagnosis;
```

```
KW   antimycobacterial; antibiotic; vaccine.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9741252-A2.
XX
PD   06-NOV-1997.
XX
PF   18-APR-1997;  97WO-EP01973.
XX
PR   29-APR-1996;  96DE-4017184.
XX
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI   Espitia C,  Honisch C,  Moreno C,  Singh M;
XX
DR   WPI; 1997-549750/50.
DR   N-PSDB; AAT93610.
XX
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacterial infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
XX
PS   Claim 5; Fig 13; 55pp; English.
XX
CC   This novel 74 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria.  Novel M. tuberculosis
CC   proteins (see AAW31851-57) are claimed.  These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals.  The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
XX
SQ   Sequence    763 AA;
```

```
alignment_scores:
       Quality:  233.00          Length:     535
         Ratio:    1.131           Gaps:      27
Percent Similarity: 38.505    Percent Identity: 25.421

alignment_block:
US-09-786-136-4 x AAW31852    ..

Align seg 1/1  to: AAW31852   from:  1  to: 763

 214 CCACAGACTGCTCCTTGGTTC......CACAGGCCATCCCGGAGGTGCT 257
     ||||:::||  ||       ||      |||::::|||
  23 ProLysSerLysProProPheProProAlaProAlaProPro.CysT 39

 258 GGCTTCGAGCCCTTCAGACAAGTAAGGTCAGCTCGCCCCCTCCTGCTTC 307
     ||:::|      |||||||||        |||||||:::||
  39 rpMet................LeuValSerAlaAlaProProCysPro 49

 308 ACTTGC.................................CC 315
     :::
  50 ProAlaProProAlaProProLysProLysSerLysAlaProPheProPr 66

 316 TGTAGCTGTGGATGCTACTACTAGGGAGGTCCTGCCT............. 352
     ||||        ::;||||||:::||
  66 oValProProAlaProProAlaArgGluLeuAlaProProLeuProProA 83

 353 .............................GGGGCCCTGCAGGTGTGC.... 370
     |:::|                        ||||||        |||
  83 laProProGluAlaProArgGluSerArgProAlaLeuProProCysPro 99

 371 .AGCTCAGACGCCGCTGAGCCCCCAGG 396
     ||||||||||       |||||||
```

```
100 ProProProValValIleProAspProProGluProAlaAlaProPro.. 115
397 AACCCCTCCTGGCCTCCACAGC.................................... 418
    :::|||||||||  ::||||
116 .ValProProAlaProAsnSerProProPheProProPheProProAlaP 132
419 ....CATCTAGATGCAGCTCCTCTGCCCACTGTTGTCTACTTCTAAAGGA 463
    :::     ||| ||:|||::|| |||        |||
132 roLysPheValProAlaProProValProProValProAsnSerProPro 148
464 CTTCAGAGAGGCCTCCAGCAGGCGCCTGGGACTCGGACCAAAATGGCAA 513
             ||||||
149 PheProPheProProAla.................................... 155
    ::
514 CTCCAAGCCTGCTTTGGGGACCCTGCCACTCCCACGGAAGGTCCTCCCC 563
    |||||||:: |||||  |||::
156 .....AlaLeuAsnProProAlaProProAlaProProLeuProLeuAlaA 169
564 GCCCACCTCCCGTCCTCCCTGT.................................... 586
    :: ||||||  |||||||:::
169 snSerProProLeuProProAlaProProThrProAlaGlyThrProPro 185
587 ....CGGCTCAGCATGGCCGCCGTCACAAACTCTGTAG 621
    :::      |||||| ::: |||  :::||
186 AlaAlaProTrpProProValProAlaAlaProLysSerLysProAlaSe 202
622 CCCTGACCCGGGCCAGGCCAACACAGTGAAGGCAGCCACCATGACTACC 671
    ||||    |||  :::||:  :::::
202 rProProArgProProAlaProProMetProAlaThrProMetGluPheP 219
672 TGCCCTTG............GTGCGGCTGCGAGGAGGCACCA 700
    |||||||         |||    |||
219 roProLeuProProValProProAspProIleSerLysGluThrProPro 235
701 GGCTCCTTCCGCCTGGACGCGCCCTTCTGCGCCGCTGTGCGCATCTCGCA 750
    |||||  ||| ||| ||||||:: |||||||
236 .AlaProProAlaProIleProProAlaProVal.................... 247
751 GGAGCGCCTCTGCCGCTGCCTTTGCCGTGCACCGGCGGCCAGCCTCA 800
    ||||||  ||||||||  ::|||  :::
248 ..ProIleProProValPro...ProLeuPro...ProValProAsnLys 261
801 GCCCACCTCGGCCTCATTGCCTGGCACTTCTGGGCCTGTGTTGGC 850
    |||||||  |||  ::|||| |||||||||||||::
262 IleProProAlaProProAlaProProValAlaValAlaAlaValLeuVa 278
851 CAGGGTGACAGTGCCACGGCCTCCTGCAGCCCGT.................C 885
    |||  |||||| :::
278 lAlaProCysProProLeuProProLeuProAsnAsnHisProProAlaP 295
886 CCCCAGCTGCGGGCTCTGAGGGGCCAGGCCAGTGGACAGTGGGCGGGGCT 935
    ||||||||| ::: ||||||  ::::::
295 roProAlaAlaProValProGlyValProLeuProLeuProAsnSer 311
936 CAGACACCGAGGCCTCCGATGGGGCGGAAGGGCTGGCGCACCGACCTG 985
    :::  ||||  |||||  ::|||||:::
312 HisProProAlaProPro...............SerAlaProValPr 322
986 CGGGGCCGGACCTGGGCCACTGCCGTAGCAC.................... 1016
    ||||::  ::|||||||||
322 oGlyValProLeuAlaProLeuProIleSerGlyArgProValSerValT 339
1016 ..................................................... 1016
339 rpLysGlySerPheThrThrLeuSerThrPheCysCysArgValCysSer 355
1017 ...................TCGCCTGGCTGGAGCACCGATGCGCC 1042
     ::|||||||:::
356 GlyGluValAlaGlyAlaLeuAsnProSerArgProSerArgSerPr 372
1043 GCTGCCTTCGACGAGTGGGAACTGACACGGCGGCCAAGGCTGATTGCTGCT 1092
     ||||  ||| ||| |||
372 oLeuThrThrThrThr.........ProAlaLeuProAlaProI 384
```

```
1093 GCGGGCCCAGCACTGCTTGCTGACGGCCTG...............ACCTGGCCG 1130
     |||  ||| |||
384 leProProLeuProProLeuProProLeuProProIleAsnThrAlaValPro 400
1131 CCCTCAAGGCCGCAGCCCGAGGGCT........................CTTCCTG 1162
     |||  |||  ||| :::  |||||
401 ProIleProProLeuProPro.ValThrAlaLeuAlaProProLeuProP 417
1163 CTACTGGCGCCACTGGGACACAAACCTGCAGCTACACCTGCT...GTGCTA 1209
     ||||||||| ::||::|||::::::: ||||||   :::
417 roLeuAlaProLeuProIleSerProGlyValProProAlaProProIle 433
1210 CAGCCCAGCGAACGTGTGAAGGCTGCCCCCTGCTGGGCTGGCGCCC 1259
     ||||:::: ::: |||||  ||| |||
434 ProProGlyLysProTrpProThrThr.ProProLeuAlaProAlaProProGly 450
1260 CACCCAACACACTCAAGTCACTGCCGCCCAGGCCTGCCTCTTGGTGCTG 1309
     |||::|||::: ||||||||
450 luProLysThrValProValLeuProProGly.................... 460
1310 GGAAAGTGTAGGCTGGTGCCAGCCTGTCCCCCACTG...CTTCTTACTCC 1356
     |||::::||||||| ||| ::||
461 ...........................ProSerCysProSerGluLysProAsnPr 471
1357 CTCCCTGGAGCCCTCTTGCCCCACCAAAAAGTGCCTGCCTGTGCTCCTCTC 1406
     |   ||::::||| ||| |||::||||:::
471 oProAlaProProGluProProGluProLysSerProAlaLeuProP 488
1407 CCTCTCCTCCCACCCACTCACACTCCCCTCCCATCCTCTGAGCTCCC 1453
     ||:::||||||:::||| :::|||||| ::: ::|||
488 roAlaProProAlaPro...SerMetProSerAlaValArgValPro 502
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW31855

seq_documentation_block:
```
ID   AAW31855 standard; Protein; 572 AA.
XX
AC   AAW31855;
XX
DT   27-APR-1998  (first entry)
XX
DE   Mycobacterium tuberculosis 55 kDa protein.
XX
KW   Tuberculosis; mycobacteria; infection; diagnosis;
KW   antimycobacterial; antibiotic; vaccine.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9741252-A2.
XX
PD   06-NOV-1997.
XX
PF   18-APR-1997;   97WO-EP01973.
XX
PR   29-APR-1996;   96DE-4017184.
XX
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI   Espitia C,  Honisch C,  Moreno C,  Singh M;
XX
DR   WPI; 1997-549750/50.
DR   N-PSDB; AAT93610.
XX
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacteria infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
XX
PS   Claim 11; Fig 16; 55pp; English.
XX
CC   This novel 55 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberulosis DNA fragment (see AAT93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
```

```
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC   proteins (see AAW31851-57) are claimed. These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals. The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
XX
SQ   Sequence   572 AA;

alignment_scores:
     Quality:   227.50        Length:         465
     Ratio:       1.132        Gaps:            30
Percent Similarity: 43.226    Percent Identity: 29.032

alignment_block:
US-09-786-136-4 x AAW31855

n seg 1/1 to: AAW31855   from: 1  to: 572

 163 ACACTGCAATTTTTCTATGAGCCCTCGTAGAGGTGGGTGGTGGGGGC 212
     ||||||:||  ||     ::|||  |||
 154 ThrLeuSerThrPheCysCysArgValCysSerGly............. 165

 213 TCCACAGACTGCTCCTTGGTTCCACAGGCCATGCCCGAGGTGCTGGCTT 262
     |||    ||||||:: |:::||||   ||||:||
 166 ..................GluValLeuAlaG 170

 263 CGAGCCCTTCAGACAAGTAAGGTCAGCTCTGCCCCCTCCTGCTTCACTTG 312
     :: ||||||:  |||::| ||| |  ||:|||
 170 ly.AlaLeuAsnProSerArgProSerArgSerProLeuThrThrThrTh 186

 313 CCCTGTAGCTGTGGATGCTACTACTAGGGAGGTCCTGCCT...GGGGCCC 359
     ||   ||::|  |||  :::  |||           |
 186 rPro...AlaLeuProAlaProIleProLeuProProLeuProProL 202

 360 TGCAGGTGTGCAGTCAGAGCCCGCTGAGCCCCCAGGAACCCTCCTGCC 409
     ||  ::    |:::::||   |||   ||||||| |||:::|
 202 euProIleAsnThrAlaValProProIleProPro...LeuProProVal 217

 410 TCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGTGTCTACTCTAA 459
     ||||||  ::|||||||||  |||   :::| |||||
 218 ThrAlaLeuAlaProProLeuProProLeuAlaProLeuProIleSerPr 234

 460 AGGACTTCAGAGAGGCCTCCCAGCAGGCGCCTGGGACTCGGACCAAAATG 509
     ||:|::: |||  |||  ||| |||:| |||||
 234 oGlyValProProAlaProProIle.ProProGlyLysPro......... 247

 510 GCAACTCCAAGCGTGCTTGGGGACCCTGCCACTC...CCACGGAAGGT 556
     |||  ||| |||| |||   ||| ::| |||||    ||:|||:|
  48 ..........TrpThrThrProProLeuAlaProAlaProPr 258

 557 CCTCGCGCCGCCACCTCCCGTCCCCGTCCTCCGCGGCTCAGCATGGGCCGCCG 606
     :::||||||||||||| |||   |||  ::|||
 258 oGluProLysThrValProValLeuProProGlyProSerCys.ProPro 274

 607 TCACAAACTCTGTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCA 656
     ||||:: ||| ||  |||      ::: :::::|
 275 SerGluLysProAsnProProLeuProProGluProProGluProLysSe 291

 657 GCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGGCTCC 706
     ||| ||| ||||||:|| ||  |  |||| |||
 291 rSerProAlaLeuPro.ProAlaProAlaProSerMetProSerAla 307

 707 TTCCGCCTG..........GACGGCGCCCTTCGGCCGCCGTGCGCAT 744
     |||::           ||| ||||| :::|||
 308 ValArgValProSerProProIle.ProProAlaProProAlaAlaP 324

 745 CTCGCAGGAGCGCGCCCTCTGCGCCGTGCCCGCCCTTTGCCGTGCACCGCGCCA 794
     ||| ||| |||   ||| :: |||


 324 roargAlaSerMetProAlaLeuPro...ProAlaProProSerProPro 339
     ::|||||||| |||:::
 795 GCCTCA................GCCCACCTCGGGCTCATTG 820
     |||::|                :::|||||| |||:::
 340 AlaThrArgLeuCysProLeuProProSerProProAlaProAsnSe 356

 821 CCCTGGGCACTTCTGGGCCCTGGTGTTGGCCAGGGTGACAGTGCCACGGC 870
     |||                     |||:|||||||
 356 rPro.........................ProAlaP 360

 871 CTCCTGCAGCCCGTCCCCCAGCTCGGGCTCTGAGGGGCCAGGCCAGGTGG 920
     ||||||||  :|| ::  ||
 360 roProAlaProProThrProProLysLeuLeu.............. 370

 921 ACAGTGGGGGGGCTCAGACACCGAGGCCTCCGATGGGGCGGAAGGGGGCTG 970
     :::|||::: |||  ||||||
 371 .........SerAlaAsnProProCysProValProProAlaProAs 384

 971 GGGGGCACCGACCTGCGGGGGCCGGACCTGGGCCACTGCCGTAGCACTTCGC 1020
     |||::|       |||  ||:::||
 384 nArgProProAlaProProAlaProProAlaProGluLeuProAlaP 401

1021 CTGGCTGGAGCACCGATGGCCGCCTGCCTTCGACGAGTGGAACTGACAG 1070
     ||  |||       |||  |||:||| |||
 401 ro........ProAspProProThrProProValAlaAsnSerProPro 414

1071 CGGCCAAGGCTGATTGCTGCTGCGGGCCCAGCACTTGCCTGACGGCCTT 1120
     |||       ::::|||    ::::||||
 415 AlaProPro.....AlaProProAlaProProSerAlaLeuProPheVa 429

1121 GACCTGGCGCCGCCCTCAAGGCCGAGCCCGAGGGCTCTTCCTGCTACTGCG 1170
     :::   ||||||||  |||||||     ::: |||
 429 lAsn...ProProAlaProPro.ThrPro...AlaAlaProLysSerArg 443

1171 CCACTGGGACCAAAACCTGCAGCTACACCTGCTGCTCACAGCCCAGCGA 1220
     |||       ||||||||| |||||| |||
 444 Pro.....AlaLeuProAlaAlaProProAlaProProAlaProProVa 458

1221 AGTGTGAAGGCTGCCCCCGCTGCTTGGCTGGCGCCCCA...CCCAAC 1267
     |||:::  ||||| |||||   :::  |||||| ||||||
 458 lArgAlaThrThr.ProProAlaProProAlaProProAlaProAsn 474

1268 ACACTCAAGTCACTGCCCGCCCAGGCGGCTGGCCTCTTGGTGCTGGGAAAGTG 1317
     |||||   ::|||||||||
 475 SerMet...AlaLeuProPro.......................... 480

1318 TAGGCTGGTGGCCAGCCTGTCCC...........CCACTGCTTCTTACTCCCTCCC 1361
     :::||| |||||||||| |||     |||||||| |||||||
 481 .......AlaProProAspProProIleProLeuLeuAlaThrProProA 495

1362 TGGAGCCCCTTCGCCCCCACAAAAAGTGCCTGCCTGTGCTCTCCCCTCT 1411
     ||||:::  |||              |||||:: ::| |||
 495 laProProAlaProPro..........LeuProMetSerProProAla 507

1412 CCTCCCACCCCACTCACACTCCCCTCCATCCTCTGAGCTCCCTGCAACAC 1461
     |||||||||| ||   |||   ::: :::||||||
 508 ProProLeuPro........ProAla.AlaProAspProProAlaProP 521

1462 AGTGGAAGGTAGAGAGCCACAGTCCCCA 1490
     ||:::||| ||||:::: |||||
 521 roLeuThrIleAsnGlnProProSerPro 530

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:ABB11413

seq_documentation_block:
ID   ABB11413 standard; peptide; 441 AA.
XX
AC   ABB11413;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human extensin homologue, SEQ ID NO:1783.
XX
```

```
CC    screening techniques. The present sequence represents a novel human
CC    polypeptide of the invention.
XX
SQ    Sequence   441 AA;

alignment_scores:
          Quality:   227.00          Length:      403
          Ratio:     1.207           Gaps:         26
Percent Similarity: 46.650    Percent Identity: 29.280

alignment_block:
US-09-786-136-4 x ABB11413

Align seg 1/1  to: ABB11413   from: 1  to: 441

348 TGCCTGGGGCCCTGCAGGTGCAGCTCAGCGCCGTGAGCCCCAGGA 397
    ||||||::||||||                     |||||    :::::
  1 CysLeuAlaProCys...............SerProGlnProGluLysAs  12

398 ACCCCTCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGT 447
    ::: ||||||               ::: |||||| ||||||
 12 nGlyMetGlnProLeu............LeuLeuLeuProProLeuL  25

448 TGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAG........GCG 488
    ||                ||:::::::                |||
 25 eu...............TyrGlnGlnLeuLeuHisSerSerLeuGlyAla  36

489 CCTGGGACTCCGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCT 538
    ||||||      ||   :::::   |||  |||| ||||
 37 ProGlyGluSerThrLeuLeuValArgThrSerLysLeuLeuValGlyLe  53

539 GCCACTC.................................CCACGGAAGGTCTCG 561
    |   |||                                 |||::::||||
 53 uGlyLeuGlnLeuLeuValTrpLeuLeuLeuGlnThrArgSerLeuLeuA  70

562 CCGCCCACCTCCCCGTCCTCCCGTCGGCTCAGCATGGGCGCCGTCACA 611
                                         ::|||
 70 laLeuGlnLeu.............................HisLeuThr  76

612 AACTCTGTAGCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGAC 661
    ::::|||::: |||      ||| |||||:::|||
 77 SerSerAlaProLeuLeu...AlaAlaProThrAlaVal..........  88

662 CATGACTACCTGCCCTTGGTCGGGCTCAGGAGGCACCAGGCTCCTTCCG 711
                         |||::||| ||| ::||||||
 89 ...............CysSerCysSerArgCysSerAlaProArgS  99

712 CC......TGGACGCGCCCTTCTGCG.................CCGCTG 737
    ::      ||||||||||||||                     |||
 99 erArgCysValAlaArgProAlaAlaArgThrGlyLeuProThrProAla 115

738 TGCGCATCTGCAGGAGCGCCTCTGCCGTGCCTGCCCTTTGCCGTGCAC 787
    ||||||       ||| :::|||  || |||
116 ProAlaSerSerProAlaAlaSerProAlaProAlaAlaSerPr 132

788 CGGGCCAGCC...TCAGCCCACCTCCGCCTCCATTGCCTGGCACTTCT 834
    |||||||||   |||:::|||  |||::|||
132 oAlaProAlaGluSerThrAlaProGlnPro................. 142

835 GGGCCCTGGTGTTGGCCAGGTGACAGTGCCACGGCCTCCTGCAGCC... 881
    ||||||::|||         |||::||||||:||||||
143 .....LeuLleLeu..........LeuProLysProProProAlaPro 153

882 ..CGTCCCCAGCTCGGGCTCTGAGGGGCCAGGTGACAGTGGG 928
    |||||| |||
154 GlyAlaProProProArgPro........................... 160

929 CGGGGCTTCAGACACCGAGGCCTTCCGATGGGGCGGGAAGGGCTGGGCGCAC 978
    |||||  |||||||
161 .GlyAlaProProArgProAla.................AlaSerP 171

979 CGACCTGCGGGGCGGGACCTGGGCCACTGCCGTAGCACTCGCCTGGCTGG 1028
```

DR   WPI; 1998-286847/25.
DR   N-PSDB; AAV62176.
XX
PT   Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT   and treatment of infection or inducing immunological response in
PT   mammal
XX
PS   Claim 10; Page 127-129; 748pp; English.
XX
CC   This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC   sequence of the invention. This sequence was isolated from a HSV-2 strain
CC   SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC   Based on homology, this sequence is a large tegument protein.
CC   The proteins can be used for the treatment or prevention of disease, to
CC   induce an immunological response in a mammal or to identify inhibitors,
CC   activators or novel antivirals. Antagonists of the proteins can be used
CC   to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC   it can also be used to induce an immunological response in a mammal.
XX
SQ   Sequence   3119 AA;

alignment_scores:
           Quality: 222.00   Length: 577
           Ratio: 0.881   Gaps: 32
Percent Similarity: 43.674   Percent Identity: 25.650

alignment_block:
US-09-786-136-4 x AAW72204

Align seg 1/1 to: AAW72204   from: 1   to: 3119   .

   15 TTTCTGGGCAGCCAGGGGCTCTTGCTCTGCTATGGGTTGAAGATTCAGC   64
      ||||:||||||:|||::::: ||| |||||||| :::||
 2439 PheValGlySerGluAlaThrGluGluLeuProPheValAlaThrHisAlaGi   2455

   65 CTGATCACTTCCTTGTCAAAGTTGGGAGTCTGAG...............C   99
      |::||||||||||||||||| :::|||
 2455 yAspValLeuAlaAlaSerAlaAlaAlaAspProPhePheAlaAargA   2472

  100 TGGTTCTGGGCCTTAGCAGGCAATCTACTCTGTTA...   134
      ||:|||||||| :::::||||||
 2472 laIleLeuGlyArgProPheAspAlaSerLeuLeuThrGlyGluLeuPhe   2488

  135 ..............................AAGTCGATAGAGGGAGAAGCTGGGAGAA 163
                                    ::: ::: |||||||| ::
 2489 ProGlyHisProValTyrGlnArgProLeuAlaAspGluAlaGlyProSe   2505

  164 CACTGCAATTTTTCTATGAGCCCTCTGTAGAGGGGTGGTGGTGGGGGCT   213
      :::   :::||| ::|||| ::||||| |||
 2505 rAlaProThrAlaAlaArgAspProArgAspLeuAlaGlyGlyAspGlyG   2522

  214 CCACA.........GACTGCTCCTTGGTTCCA.....CAGGCCATGCCC   248
      :::|| ||| ||| ||| ||
 2522 lySerGlyProGluAspProAlaAlaProProAlaArgGlnAlaAspPro   2538

  249 GGAGGTGCTGGCTTCGAGCCCTTCAGCACAAGTAAGGTCAGCTCTGCCCCC   298
      ||| |||||||| ::ProT   2544
 2539 Gly.ValLeuAla.

  299 TCCTGCTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGGAGGTCCT   348
      ::: ||||| || :||||||||||| ||| ::
 2544 hrLeuLeu.Thr.................AspAlaThrThrGlyGluProVa   2555

  349 GCCT........................GGGGCCCTGCAGGTGTGCAGCT   374
      ||| ::|||:|:|||:|||
 2555 lProProArgMetTrpIleTrpIleHisGlyLeuGluLeuAlaSerA   2572

  375 CAGAG....................CCCGTGAGCCCCCAGGAACCCCT   403
      :::||| |||||
 2572 spAspAlaGlyGlyProThrProAsnProAlaProAlaLeuLeuProPro   2588

  404 CCTGCCCTCCAGCACCATCTAGATGCA........GCTCCTCTGCCCAC   444

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT:AAW72204

seq_documentation_block:
ID   AAW72204 standard; Protein; 3119 AA.
XX
AC   AAW72204;
XX
DT   13-JAN-1999 (first entry)
XX
DE   HSV-2 strain SB5 Contig ID 15 ORF#39 protein.
XX
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor.
XX
OS   Herpes simplex virus type 2.
XX
PN   WO9820016-A1.
XX
PD   14-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US20016.
XX
PR   09-JUN-1997;   97US-0049018.
PR   04-NOV-1996;   96US-0030279.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI   Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI   Esser KM, Leary JJ;
XX

   171 ||:::   ::||||||::|||::||| ||| ||||||
       roSerProAlaSerProAlaPro...ProAlaAlaSerProValLeu   186

  1029 AGCACCGATGCGGCCGCTTCGACGAGTGGGAACTGACAGCGGCCAAG 1078
       ::::::: :|||||||::||||||:|:|| :::::|||
   187 ThrAlaSerProLeuProAlaAlaSer............ProSe   198

  1079 GCTGATTGCTGGCTGCGGGCCCAGCACTTGCTGACGGCCTTGACCTGGC 1128
       ||:::   :::||| ::::|||||
   198 rProAlaSerProAlaProProAlaProAlaSerProValLeuThrAlaS 215

  1129 CGCCCTCAAGGCCGCAGCCCGAGGGCTCTTCCTTGCTGCGCCACTGGG 1178
       ||| ||| ::: ||| |||||||::|:|::|||
   215 erProProLeuProAlaAlaSerProSer.ProAlaAlaSerProAlaPr 231

  1179 ACCAAAA.................CCTGCAGCTA 1195
       |||||| ||||||||:::
   231 oProAlaAlaSerProValLeuThrAlaSerProProLeuProAlaAlaS 248

  1196 CACCTGCTGTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCTGCTGC 1245
       ::||||||:::   :: ||||||:: |||||||
   248 erProAlaLeuAlaAlaSerProValHisThrAlaSer.ProProValHi 264

  1246 TTGGGCTGGCGCCCCACCCAACACACTCAAGTCACTGCGCGCCAGGGCTG 1295
       ||||:: ||| :|||| :: ||||||
   264 sValAlaSerProValHisThr......AlaSerProPro........   276

  1296 GCCTCTTGCTGGTGCTGGGAAAGTGTAGGCTGGTGGCCAGCCTGTCCCCACTG 1345
       :::   :::|||||| ::||||| || ||||
   277 ....ValHisValAlaSerProProValHisThrAlaSerProProVal 291

  1346 ..CTTCTTACTCCCTCCCTGGAGCCCCTTGCCCCCAAAAAGTGCCT 1392
       :::::   :::::||| ||||||
   292 HisValAlaSerProProValHisThrAlaSerProHisValHisValAl 308

  1393 G.......CCTGTGCTCTCCCCTCTCCCCACCCACTCACACTCCCCT 1436
       ||||| |||||||||:: || ||||||
   308 aSerProProValHisValAlaSerProProVal.HisValAlaSerPro 324

  1437 CCA 1439
       |||
   325 Pro 325

```
2844  roLeuAlaProGlyProValAlaProSerGluProLeuCys...GlyTrp 2859
                                            CGCCCACCCAACACACTCAAGT
1255  ..............................             ::::::  ::       1277
2860  ValValProGlyGlyProValAlaArgArgProProProGlnSerProAl 2876

1278  CACTGCCGCCCAGGGCTGGCCTCTTGGTGCTGGGAAGTGTAGGCTGGTG 1327
      :|||  ||| |||           :::             :::||| |||
2876  aThr.LysProAlaAlaArgThrArgIle......ArgAlaArgSerVal 2890

1328  CCAGCCTGTCCCCACTGCTTCTTACTCCCTCCTGGAGCCCTTCTGCC 1377
      ||  |||||||||| |||||||||              ::|||  |||
2891  ProGln...ProProLeuProGlnProProLeuProProLeuPr 2906

1378  CCACAAAAGTGCCTGCCGTGCTCTCTCCTCCTCCTCCACCCCACTCA 1427
      |:::                   ||| |||  ||||||
2906  oGln................ProProLeuProGlnProProLeuP 2916

1428  CACTCCCCTCCATCCTCTGAGCTCCCTGCAACACAGTGGAAGGTAGAGA 1477
      |||::||||||           :::  |||        ||||||
2916  ro.GlnProProLeuProProProGlnProProLeuProGl 2932

1478  GCCACAGTCCCCAAATCCT 1496
      :||||      |||:||||
2932  nProProLeuProGlnPro 2938

seq_name: /SIDS1/gcgdata/geneseqp-embl/AA2001.DAT:ABG03533

seq_documentation_block:
ID   ABG03533 standard; Protein; 900 AA.
AC
XX
AC   ABG03533;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #3524.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,   Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS67720.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 33892; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
```

```
2589  ProAlaThrAspGlnSerValProThrSerGlnTyrAlaProArgProIl 2605
      ||||||::  ::         ::                    |||
445   TGTTGTTCACTCTCTAAAGGACTTCAGAGAGGCTTCCAGCAGCGCCTGGG 494

2605  e.......................GlyProAlaAlaThrAlaArgG 2613
                     .........GlyProAlaAlaThrAlaArgG

495   ACTCGGACCAAAATGGCAACTCCAAGGTGCTTGGGGGAC...CCTGCC 541
      :::::                       ||| |||     |||:::
2613  luThrArgProSerValProProGlnGlnAsnThrGlyArgValProVal 2629

542   ACTCCCACCGAAGGTCCTCGCCGCCCACCTCCCGT.............. 577
      ::|||  ::: |||||| |||||
2630  AlaProArgAspAspProArgProSerProProThrProSerProProAl 2646

578   ............CCTCCCTGCCGTCAGCATGGGCCGCCGTCACAAAC 614
                  :::||          ||  ::::||||||||||
2646  aAspAlaAlaLeuProProProAlaPheSer.GlySerAlaAlaAlaPhe 2662

615   TCTGTAGCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCAGCGACCAT 664
      |||::::||:|| |||:::|||        :::::      :::
2663  SerAlaAlaValProArgValArgArgSerArg...ArgThrArgAlaLy 2678

665   GACTACCTGCCCTTGGTGCGGCGTCGAGGAGGCACCAGGCTCCTTCCGCCT 714
      :::  :::||| |||            |||  |||:||::||||
2678  sSerArgAlaPro..........ArgAlaSerAlaAlaProProGluGlyT 2691

715   GGACGCCGCCCTTCTGCGGCCGCTGCCCATCTCGCAGGAGCGCCTCTGCC 764
      ||||                        ::|||||||||||||
2691  rpArgProProAlaLeuProAlaProValAlaProValAlaAlaSerAla 2707

765   GTGCCTCGGCCCTTTGCGTGCACCGCCGCTCAGCCCCACCTCGGCC 814
      |||          |||  |||  |||::|||||||||||||
2708  ArgProProAspGlnProProThr...ProGluSerAlaProPro..... 2721

815   TCATTGCCTGGGACTTCTGGGCCTGTGTTGGCCAGGGTGACAGTGC 864
      |||   ||| |||           |||  |||||
2722  ..........AlaTrpValSerAlaAlaLeuProLeuProP 2731

865   CACGGCCTTCCTGCAGCCCGTCCC.....CCAGCTCGGGCTCTGAGGGGC 908
      ||  |||       ||||||||       :::|||   |||
2731  roGlyProAlaSerAlaArgGlyAlaPhe.ProAlaProThrLeu...... 2745

909   CAGGCCAGGTGGACAGTGGGCGGGGCTCAGACACCGAGGCCCTCCGATGGG 958
                 |||||    :::||  ||||||||||||
2746  ..............AlaProIleProProProPr 2752

959   GCGGAAGGGCTGGGCGGCACCG.................A 981
      ::||||   :::|||||
2752  oAlaGluGlyAlaValAlaProGlyAspAspArgArgGlyArgArgG 2769

982   CCTCGCGGGCCGGACCTGGGCCACTGCCGCTAGCACTCGCCTGCCTGGAGC 1031
      |||||||||||:::::|||       ||| ||::::|| |||::|||
2769  lnThrThrAlaGlyTyrProSerProThrProProArgGlyPro.AlaAlaGl 2785

1032  ACCGATGCCGCCGCCCGCCTTCGACCAGTGGGAACTGACAGCGGCCAAGGCT 1081
      :||| ||||||   ||       ||   ::|||
2785  yProProArgArgLeuThrArgProAla.................. 2794

1082  GATTGCTGGCTGCGGGGCCCAGCACTTGCCTGACGGCCCTTGACCTGGCCGC 1131
      ||||||     ::|||  ::|:||     |||  |||
2794  ..............ValAlaSerLeuAlaSerLeuAsnSerProSerProArgAsp 2810

1132  CCTCAAGGCCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCACTGGGACC 1181
      ::::|||:||:::  ||||||||:||:||| ||:::||||  ::  ||
2811  ProAlaAspHisAlaAlaAlaValSerAlaAlaAlaAlaAlaValProPr 2827

1182  AAAACCTGCA.........................GCTACACCTGCTG 1204
      |:::|||          ::::|||
2827  oSerProGlyLeuAlaProThrSerAlaValGlnThrSerProProP 2844

1205  TGCTACAGCCAGCGAACGTGTGAAGGCTGCCCCCTGCTGCTTGGGCTGG 1254
      |||  :::  ||| |||  |||       ||||||
```

```
CC    quantitating a polypeptide in tissue, as molecular weight markers and as
CC    a food supplement. (II) and its binding partners are useful in medical
CC    imaging of sites expressing (II). (I) and (II) are useful for treating
CC    disorders involving aberrant protein expression or biological activity.
CC    The polypeptide and polynucleotide sequences have applications in
CC    diagnostics, forensics, gene mapping, identification of mutations
CC    responsible for genetic disorders or other traits to assess biodiversity
CC    and to produce other types of data and products dependent on DNA and
CC    amino acid sequences. ABG00010-ABG30377 represent novel human
CC    diagnostic amino acid sequences of the invention.
CC    Note: The sequence data for this patent did not appear in the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence    900 AA;

alignment_scores:
       Quality:   220.50           Length:    551
       Ratio:     0.930            Gaps:      35
Percent Similarity:  43.013        Percent Identity:  27.042

ment_block:
09-786-136-4 x ABG03533       .

Align seg 1/1    to: ABG03533    from: 1  to: 900

  201 GGTGGTGGGGCTCCACAGACTCCTCCTTGGTTCCACAGGCCATGCCCGG 250
                       ||||| ||:|:||   ::::|      |||||
  177 GlyProGlyGlySerProArgCysAsnArgCysArgGluArgLysProGl 193

  251 A...GGTGCTGGCTTCGAGCCC...TTCAGACAAGTAAGGTCAGCTCTGC 294
      | ||| ||||: |||  |||||   :: :         |||  |||||||
  193 yThrGlyProGlyTrpProProArgLeuArgSerProGlyAsnLeuArgP 210

  295 CCCCTCCTGCTTCACTTGCCTGTAGCTGTGGATGCTACTACTAGGGAGG 344
      ||  |:::::|||:::||| |||
  210 roGlyValGlyLeuGlyLeuAlaLeu.......................  219

  345 TCCTGCCTGGGGCCCTGCAGGTGTGCAGCTCAGAGCCCGCTGAGCCC... 391
              |||::||||| |||||||| :|:||| ::||||
  220 ......ProAlaArgThrAlaAlaAlaAlaPro..ArgProArg       231

  392 .........CCAGGAACCCTCCTGCTCTCCACAGCCATCTAGA        426
               ||||||:::||| :|                  ::
  232 GluArgTrpArgSerProGlyAlaProCysLeuGly.......AlaGl  245

  427 TGCAGCTCCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCT 476
      :::: |||||||||:::::||| :|:||| |||          ::|||
  245 n***ProSerLeuProAlaAlaGluProArgGly.......ProGlyT  260

  477 CTCCAGCAGCCGCTGGGACTCGGACCAAAATGCAACTCCAAG...CGT 523
      ::: :: |||||| :|||:  |||||||||||| ::||:|||     |||
  260 hrSerGlyArgAlaTrpProSerSerAlaAlaAlaArgGluAspCysArg 276

  524 GCTTTGGGG...GACCCTGCCACTCCCACGGAGGTCCTCGCCGCCACC 570
      ||| |||  |||||||||||:|||||||| ||| ||||| ||||
  277 AlaAProGlyArgGlyProAlaAlaProThr...GlyAlaArgArg...Pr 291

  571 TCCCGGTCCTCCCCGTGCGGCTCACATGGGCCGCGTCACAAACTCTGTA 620
      ||||||||| ::: ||:::|
  291 OProArgProGlyAlaGlyLeuAla.......................  299

  621 GCCCTGACCCGGCCAGGCCAACAACAGTGAGGCAGCCACCATGACTAC 670
                     ||||||||| ||:|||||||
  300 ........Gly.ArgProThrArgAlaArgAla...............  307

  671 CTGCCCTTGGTGCGGCTGCGAGGAGGCACCAGCCTCCTTCCGCCTGGACGC 720
      ||||||        |||:|| ::: :|||||
  308 .LeuProGlyArgIleGlyArgSerProGlyArgAlaProHisS  324

  721 GCCCTTCTCGCCGCCGTGCCGCATCTCGCAGGAGCGCCCTCTGCCGTGCCT 770
      ||||:: |||||||| |||||| |||:|||||||
```

  324 erProAlaGlyLeuArgSerAlaAlaArg.................CysPro 336

  771 CGCCTTTCCGTGCCACCGGCCAGCCTGCCTCAGCCCACCTCGGCCTCATTG 820
      ||| ||||| ||||| ::: |||| |||:||
  337 ProProGlyProCys..AlaProArgGlyGlnProGlyLeuProArgGl 352

  821 CCCTGGG.........CACTTCTGGGCCCTGGTGTTGGCCAGGG 855
      ||| ::: |||||||||| ::|||
  352 yGlyGlyAlaAlaProLeuArgHisGlyGlyGlyAlaGlnLysThrArgP 369

  856 TGACAGTGCCACGGCCTCCTGCAGCCCGTCCCCAGCTCGGGCT....... 899
      ::: |||||| ||||| |||||||||||||
  369 roSerGlnGlyArgProGlyProProProProAlaProAlaGlyGlu 385

  900 ....CT 901
  386 ArgGlnProGlnProGlyLysGluSerGlyGluLysGlyValGlnInProAl 402

  902 GAGGGGCCAGGCCAGGTGGACAGTGGGCGGGGCTCAGACACCGAGGGCCTC 951
      ||||||::: ||| ||| ||| :::::|||
  402 aArgGlyGlu..ProGlnGlyProGlnArgAlaMetArg............. 414

  952 CGATGGGGCGGAAGGGCTGGGCCGCACCGACCTGCGGGGCCGGACCTGGG 1001
      ||||: |||| ::||| :::| |||
  415 ......GlyAlaProAlaGlyArgArgProArgProProAlaProAlaProGl 429

 1002 CCACTGCCGTAGCACTCCGCTGGCTGGAGCACCGATGCGCCGCTGCCTTC 1051
      ::::: ||||||| :::|| ||||||
  429].................AlaAlaAlaGlnPro.................G 435

 1052 GACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGGCTGCGGGGCCCA 1101
      ||| ||| ||||||||||||
  435 lyArgTrpGlyThrAlaAlaGly.......AlaAlaGlyPro   446

 1102 GCACTTGCCTGACGGCCTTGACCTGGCCGCCCTCAAGGCCGCAGCCCCAG 1151
      |||||| |||||| ||||||||| |||||| ::||||
  447 Ala...AlaGlyArgProAlaProGlyArgGlnProSerArgAlaAlaGlyl 462

 1152 GGCTCTTCCTGCTACTGCGCCA...................... 1173
      :::: ||||||:::||||||
  462 uGly...ProAlaAlaAlaProGlyGluArgGlyGlnSerIleProValP 478

 1174 ...............................CTGGGACCAAAAA 1185
                                      |||||
  478 rolleGlyValValGluGluArgGlyProGlyProGlyProThr 494

 1186 CCTTGCAGCTACACCTGCTGTGCTACAGCCCAGCGAACCGTGTGAAGGCTGC 1235
      |||| ||| ::||||||||||
  495 AlaAlaAlaGlyProAlaAlaProGlyArgArgGlyArgArgGlySerAl 511

 1236 C..........................CCCTGCTGTTGGGCTGGCGCCCCAAC 1267
      |:::   ::||||| :::|||
  511 aArgAlaArgGlyProGluProAlaAlaProAlaArgGlnProGlyArgA 528

 1268 ACACTCAAGTCACTGCCGCCCAGGGC........................ 1293
      :::::::::|| ||||||:||:|
  528 rgAlaArgLeuProArgThrArgGlyProProArgGlyProAlaProSerGlu 544

 1294 ...TGGGCTCTT...............GGTGCTGGGAAAGTGTAGGCTGG 1325
      ||||||||||| :::::::||:||| :::||||
  545 ProTrpProLeuProSerArgProProGlySerGlyLys......ArgGl 559

 1326 TGCCAGCCTGTCCCC.................ACTGCTTCTTACTTCCC 1357
      ||||||| :::::::::|||||||| :::::||| :::|||
  559 yAlaSerAlaProGlyProGlyCysProSerArgSerGlyAsn..SerPro 575

 1358 TCCCTGGAGCCCCTTCTTGCCCCCACCAAAAAAGTGCCTGCCTGTGCTCTCTCC 1407
      ||::||:: |||||||:::|| :|||| :||||||
  576 ThrLeuGlnAspProGlyProHisSerGlyCys...ProGlu.GlyPheP 591

 1408 CTCTCCCCCAACCCCACTCACACTCC..................... 1433
      |||:|||::::|||||||||
  591 roArgLeuAlaGlyGlyGlyGluArgSerGlyProGluMetGlyThrVal 607

```
1433 ..............................................  1433
 608 GlySerAsnAspValValGlyAsnArgGlyLysThr***GlySerSerTr  624
1434 ......................CCTCCATCCTCTGAGCTCC  1452
                          |||||||||:::|||
 624 pValGlnArgThrTrpAlaGlyGluLeuProProGlySerMetGlyA  641
1453 CTGCAACACAGTGGAAGGGTAGAGAGCCACAGTCCCAAATCCT  1496
     :::|||||||||   |||:||||:::|||
 641 laGlyThrGluTrpTyrGlnGlyThrProGluAsnProLysPro  655

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:AAY50558

documentation_block:
AAY50558 standard: Protein; 419 AA.

XX
AC   AAY50558;
XX
DT   25-JAN-2000 (first entry)
XX
DE   HSV2 LAT protein.
XX
KW   LAT; latency associated transcript; regulatory; drug; identification;
KW   infection.
XX
OS   Herpes simplex virus.
XX
PN   WO9953043-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999; 99WO-GB01111.
XX
PR   09-APR-1998; 98GB-0007865.
XX
PA   (UNLO ) UNIV COLLEGE LONDON.
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Coffin RS, Latchman DS, Thomas SK;
XX
DR   WPI; 1999-633828/54.
DR   N-PSDB: AA223924.
XX
PT   New polynucleotides used to identify drugs which prevent reactivation
     of latent Herpes simplex virus (HSV), treating HSV infections -

Disclosure; Page 35-37; 37pp; English.

CC   This invention describes a novel isolated polynucleotide (I) essentially
CC   having an open reading frame (ORF) encoded by a herpes simplex virus
CC   (HSV) latency-associated transcript (LAT) operably linked to a
CC   regulatory sequence permitting expression of a LAT ORF polypeptide in a
CC   cell. The polypeptides are used to identify drugs that prevent
CC   reactivation of latent HSV for the treatment of HSV infections in
CC   mammals, particularly humans. This sequence represents the Herpes Simplex
CC   Virus type 2 LAT protein described in the invention.
XX
SQ   Sequence  419 AA;

alignment_scores:
        Quality:  220.00      Length:   494
          Ratio:    1.146       Gaps:    30
Percent Similarity:  38.866   Percent Identity:  26.113

alignment_block:
US-09-786-136-4 x AAY50558   ..

Align seg 1/1  to: AAY50558  from: 1  to: 419

 104 TCTGGGCCTAGCAAGGCAATCTACTCTTGTTAAAGTCGATAGAGGGAGAAG  153
```

```
                       |||::: |||:::::||   :::::|||          |||||
   6 SerSerCysSerSerPheLeuProAlaArgValArgValArgGlyAr  22
 154 CTGGGGAGAACACTGCAATTTTTCATGAGCCCTCTGTAGAGGGTGGGT  203
                       |||:::     |||::
  22 gGlyGlyInValCysGly..........................ArgGlyGly.  31
 204 GGTGGGGGCTCCACAGACTGCTCCTTGGTTCCACAGCCATGCCCGAGG  253
                       |||:::         |||:||||  |||||| |||||
  32 ...........ArgArgGlyGlyGlyHisGlyArgLys  41
 254 TGCTGGCTTCGAGCCCTTCAGACAAGTAAGGTCAGCTCTGCCCCTCTG  303
                       |||   |||::|||   |||::||||:  |||::|||||
  42 CysGlyCys...........MetProArgValPheSerProSerSe  54
 304 CTTCACTTGCCCTGTAGCTGTGGATGCTACTACTAGGAGGTCCTGCCTG  353
                       |||   ||||||                      .........ProG  61
  54 rPheLeuProProVal...........
 354 GGGCC.......CTGCAGGTGTGCAGCTCAGAGCCCGCTGAGCCCCCA  394
                       |||:::     |||    :: |||::  ||||||
  61 lyGlyArgGlyAlaCysValValIleGInProProSerProProPro  77
 395 GGAACCCCTCCTGCC.........................................  409
                       ||||||:::
  78 HisPheProProSerLeuSerLysPheProGlyProTrpLeuArgArgAlaGl  94
 410 ........................................................TCTC  413
                                                              |||
  94 yGlyAlaAlaAspProProProProSerProSerGlnAlaProSerH  111
 414 ACAGC..................................................  418
     |||||
 111 isSerCysProProProGlySerGlyArgProAspValArgAlaLeu  127
 419 CATCTAGATGCAGCTCCTCTGCC............ACTGTGTCTAC.  454
     ||| ||| |||||||||||||||                |||::: :::
 128 His..AspArgAlaProLeuProValAsnThrAspThrLeuPhePheTy  143
 455 ........................................TCTAAAGGACTTCAGAGAGGC.  475
                                              ::||||||
 143 rProProAlaSerProProThrHisGlnAspArgGluProGluArgGlyA  160
 476 .TCTCCAGCAGCGGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGT  523
     :::|||||        |||||||||||||::|||||||||||
 160 rgAlaProAlaLeuPheTyrAspLysAspGlnGlnAlaSerGlyValGly  176
 524 GCTTTGGGGGACCCTGCACTCCACGGAAGGTCCTGCCGCCACCTCC  573
     |||  :::          |||::::::::          ||| ||
 177 AlaAlaSerArgAlaArgProProSerSerSerLeuProProHisPr  193
 574 CCGTGCTCCCCTGTCGGCTCAGCATGGGCCGCCGTCACAAACTCTGTAGCC  623
     ||||||||||||                |||:: ||| :::
 193 oArgProProCys.........AlaGlyGluLeuHis.Gln...........  203
 624 CTGACCCGGGCCAGGCCAACAACA.......GTGAAGGCAGGACCATGAC  667
     ||||||||||||               ||||||||:
 204 ............ArgProThrThrLysCysValLysSerIleThrLysLe  216
 668 TACCTGCCCT........................................T  678
                                                        |||
 216 uTyrCysLysIlePheIleAsnIleLysPhePheSerSerSerPheG  233
 679 GGTGCGGCTGCAGGAGGCACCAG...................GCTCCTTCC  710
                       |||   :::  |||:::
 233 lnGlnGlyGlnInLysValHisAsnLysMetLeuValCysValAlaValArg  249
 711 GCCTGGACGCGCCCTTCTGCGCCGCTGTGCGCCATCTCGCAGGAGCGCCTC  760
     :::  :::||||| |||||
 250 GlyArgValArgProProProLeu.........................  258
 761 TGCCGTGCTGCCGCCCTTTGCCGTGCACCGCGCCAGCCTCAGCCCCACCTC  810
                       ||| |||||||||                       |||||||
```

```
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI: 2001-639362/73.
DR   N-PSDB: AAS67720.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 33892; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note:The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   900 AA;

alignment_scores:
        Quality:   218.50                Length:     593
          Ratio:    0.892                 Gaps:        38
Percent Similarity:   41.315     Percent Identity: 26.813

alignment_block:
US-09-786-136-4/rev x ABG03533   .

Align seg 1/1   to: ABG03533   from: 1   to: 900

1411 AGAGGGAGAGAGCACAGGCAGGCACTTTTGTGGGGCAAGAGGGCTCCA 1362
         |||   :::::||||||||
  29 ArgProAlaAlaGlnLysGlnAlaLeu.................. 37

1361 GGGAGGGAGTAAGAAGCAGTGGGGGACAGGCTGGCACCAGCCTACACTTT 1312
         ||||||||||  :::::||  ||||||||
  38 ....GlySerArgGluArgValGlyThrGly.............. 46

1311 CCCAGCACCCAAGAGGCCAGCCCTGGGCCGGCAGTGACTTGAGTGTGTTGGG 1262
         ||||||||    :::|||||:   |||
  47 ..................ProGlyArgIle...LeuArgProGlyGly 55

1261 TGGGGCGGCCAGCCCAAGCAGCAGGGGCAGCCTTCACACGTTCGCTGGGC 1212
         ||||||:::  |||||:  |||||||    ||
  56 TrpGlyCysPheProGlyProArgGlyThrGluAspAlaArgGlnArgAl 72

1211 TGTAGCACAGCAGTGTAGCTGCAGGTTT............... 1182
         ||||||||:::||||||::||||
  72 a..AlaArgGlyProValGlyAlaGlyThrGlnHisGlyArgAlaAlaVa 88

1181 .................................GGTC 1178
                                           |||
  88 lProArgGlyProGlnAsnGluProAspGluThrLeuLeuProGlyGlyP 105

1177 CCAGT.........GGCCGCAGTAGCAGGAAGAGCC.......... 1152
         |||||         |||   :::  |||||||||:::
```

```
259 ......ProProProLeuProValSerSerProSerPheProProProT 273
811 GGGCTCATTGCCCT.........................GGCCACTTC 833
        ||||||                                  |||
273 hrSerProCysProArgGlyArgAlaSerAlaGlyGlyProValGlyGlyGly 289
834 TGGGCCCTGGTGTTGGCCAGGGTGACAGTGCCA................ 866
        :::  |||||||  ::::||||  :::|||
290 PheLeuArgAlaAlaSerArgValLeuAlaProProThrProArgGlyPr 306
867 ...............CGGCCTCCTGCAGCCCGTCCCCCAGCTCGGCGTCTGA 903
                       ||||||      ::::::
306 oArgGlyArgArgArgProAlaArgProAlaProGluThrThrG 323
904 GGGGCCAGGCCAGGTGGCAGTGGCGGGGGCTCAGACACCGAGGCCTCCG 953
        :::::||||||     ::||| |||
323 lyGly..........AlaThrGlyGlyArgGlySerProAlaArgPro 335
954 ATGGGGCGGGAAGGGCTGGGCGGCACCGACCTGCGGGCCGGACCTGGGCC 1003
        ||    ||  |||||  |||||||||
336 ProGlyCysProGlyGlnArgAla...CysCysProProGly........ 348
004 ACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGCCGCTGCCTTCGA 1053
        |||     ||||||  ||||||       ::::
349 ...........SerSerArgGlyCysSerThr.....ProThrArgGlyA 360
1054 CGAGTGGGAACTGACAGCCGCCAAGGCTGATT...............GCT 1088
        ::||||||::   |||                              |||
360 laSerGlySerArgGlyAlaSerArgSerCysAlaAlaArgAlaArgGlyAla 376
1089 GGCTGCGGGCCCAGCACTTGCCCGACGCGGCTTGACCTGCCGCCCTCAAG 1138
        ||||||  |||:::::||||||
377 GlyCysGlySerThr......................ArgGl 384
1139 GCCCAGCCCGAGGGCTCTTCCTGCTCACTGCGCCACTGGGACCAAAACCT 1188
        |||:::   |||  |||    :::||| |||||:::
384 yProArg..ProAlaAlaGlyProProArgSerProSerGlyProArgGly 400
1189 GCAGCTACACCTGCTTGCTGTGCTACAGCCCAGCGAACGTGCAAGGCTGCCCC 1238
        |||  ||||||| |||||  ::||||:::
401 ProArgArgPro.LeuCys...GlyProSer................ProC 411
1239 CTGCTGCTTGGGCTTGGCGCCCCCACCC 1264
        :::  ||||||||||::::||||||
411 ysProGlyArgAlaGlySerProPro 419

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03533

seq_documentation_block:
ID   ABG03533 standard; Protein; 900 AA.
XX
     ABG03533;
XX
                              -3-FEB-2002  (first entry)
DE   Novel human diagnostic protein #3524.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
```

```
 105 roSerProArgGlyGluLeuArgGlyArgSerGlyAlaArgGlyLeu 121

1151 ..........CTCGGGCTGCGCCCTTGAGGCGCGCAGGTCAAGGCCG 1114
                |||    |||     :::    |||
 122 Pro***SerLeuThrGlyProAlaProGlyProGlnArgGlyGlyGly** 138

1113 TCAGGCAAGTGCTGGGCCGCAGCCAGCAATCAG...CCTTGGCCGCTGT 1067
     :::   |||  |||::::|           |||||:||||
 138 *SerProSerProGlyArgAlaSerSerLysAlaGlyLyProTrpLysArgP 155

1066 CAGTTCCCACTCGTCGAAGG...CAGCGGCCCATCGGTGCTCCAGCCAGG 1020
     ::|||    |||||||        :::  ||||||
 155 roGlyAlaSerArgAlaSerLeuGlnArgAlaSerSerMetProAlaSer 171

1019 CGAGTGCTACGGCAGTGGCCCAGGTCCGGCCCAGCAGT.......... 981
     ::|||    :::|||    :||||||:|||
 172 GlnVal...AspTrp......GlyGlyProGlyGlySerProArgCy 184

 980 ...CGGTGCCGC...........CCAGCCTTCCGCCCCATCGGAGGCCTC 944
     ||||||||   :::::||||||||       |||
 184 sAsnArgCysArgGluArgLysProGlyThrGlyThrGlyTrpProProA 201

 943 GGTGTCTGCAGCCCCCGCCCACTGTCCACCTGCCTGGCCCCCTCAGAGCCCG 894
     ||  |||||||:::|||   ||||||||:||:|:
 201 rgLeuArgSerProGlyAsnLeuArgProGlyValGly............. 213

 893 AGCTGGGGGACGGGCTGCAGGAGGCCG...................... 867
     ||| |||  |||
 214 ........GlyLeuGlyLeuAlaLeuProAlaLeuProAlaThrAlaAlaAlaAlaAlaPr 228

 866 ........TGGCACTGTCACCCTGGCCAA.................C 845
     :::|||||||:||
 228 oArgProArgGluArgTrpArg...SerProGlyAlaProCysLeuGlyA 244

 844 ACCAGGGCCCAGAAGTGCCCAGGCAATGAGAGCCGAGGTGG......... 804
     ||||||||||:|||||||  |||||
 244 laGln***ProSerLeuProAlaAlaIleGluProArgGlyProGlyThr 260

 803 ..GGCTGAGGGTGG.................................... 792
     ||| :::|||
 261 SerGlyArgAlaTrpProSerSerAlaAlaAlaArgGluAspCysArgAl 277

 791 ........CCGGGTGCACGGCAAAGGGCGAGGC 767
     |||||||||:||| :::
 277 aProGlyArgGlyProAlaAlaAlaProThrGlyAlaAlaArgArgProProArgP 294

 766 ACGGCAGAGGCGCTCCTGCGAGATGCCGCACAGCGGCCACGGCAGAGGGCCCGT 717
     ||||| |||     ||||||
 294 roGlyAlaGlyArg......ProThrArgAlaArgAlaLeu 308

 716 CCAGGC.....GGAAGGAGCCCTGGTGCCTCCTGCAGCGCGCACCAAGG.. 675
     ||||||       ||| ||||||||||:::
 309 ProGlyArgIleGlyArgArg...SerProGlyArgAlaAlaProHisSe 324

 674 ....GCAGGTAGTCATGTCGCTGCCTTCACTGTTGGCCTGGCCCCGG 629
     |||||||  :::   ||||||||||
 324 rProAlaGlyLeuArgSerAlaAlaArgArgCysSerProProGlyPro. 340

 628 GTCAGGGCTACAGAGTTTGTGA.......CGGCCGGCCCATGCTGAGCCGA 586
     |||          :::|||  |||  |||
 341 ...............CysAlaProArgGlyGlnProGlyLeuProArg 351

 585 CAGGGAGGA..........CGGGGAGGTGGG.....CGGCGAGG 557
     ||||||          ::|||||||       :::
 352 GlyLyGlyAlaAlaAlaProLeuArgProHisGlyGlyValAlaGlnLysThrAr 368

 556 ACCTTCCGTGGGAGTGGCAGGGTCCCCAAAGCACGCTTGGAGTTGCCAT 507
     |||||||  :::  ||||||||||:::
 368 gProSerGlnGlyArgProGlyProPro.................... 377

 506 TTTGGTCCGAGTCCCAGCGCCGCTGCGAGAGCCTCTCGAAGTCCTTTA 457
     ||||||||||||||||||
 378 ........ProProProAlaProAlaGlyLyGluArgGlnProGlnPro... 390
```

```
 456 GAGTAGACAACAGTGGCAGAGGAGCTGCATCTAGATGGTCTGAGAGGC 407
     |||   |||::::::::::  :::::::::::  ||
 391 ..........GGLysGluSerGlyLuSerGlyGlyValGlnProAl 402

 406 AGGAGGGTTCCT...GGGGGCTCAGCGGGCTCTGAGCTGCACACCTGCA 360
     | |||  |||  |||||||:||      :::||||||||
 402 aArgGlyGluProGlnGly.ProGlnArgAlaMetArgGlyAlaProAla 418

 359 GGG........CCCAGGCAGGACCTCCCAGTAGTAGCATCCACAGCT 319
     |||       ||||||||||::::::
 419 GlyArgArgArgProProAla.ProAlaProGlyAlaAlaAlaGlnProG 435

 318 ACAGGGCAAGTGAAGCAGGAGGAGGGGCAGAGCTGACCTTACTTGTCTGAAG 269
     :: ::: :::::|:|||||:::
 435 lyArgTrpGlyThrAlaAlaGlyTyrAlaAlaGlyPro..........Ala 447

 268 GGCTCGAAGCCAGCACCT.........................CC 249
     :::::::::||||||||||                         ||
 448 AlaGlyArgProAlaProGlyArgGlnProSerArgAlaAlaAlaGluGlyPr 464

 248 GGGCATGGCCTGTGGACCAAGGAGCAGTCTGTGGAGCCCCCACCACCCA 199
     |||   :::||||||||||||      || :::|||||:
 464 oAlaAlaAlaProGlyGluArgGlyGlnSerIleProValPro....... 478

 198 CCCTCTCACAGAGGGCTCATAGAAAA.................... 173
     |||||||||||||:::
 479 .......IleGlyValValGluLuGluArgGluGlyProArgGlyProGly 492

 172 ..ATTGCAGTGTTCTCCCCAGTTCTCCCTCTCATCGACTTAACAGAGT 126
     :::|||||||||||::|||
 493 ProThrAlaAlaAlaGlyProAlaAlaPro..........GlyArgGl 505

 125 AGATTGCCTTGCTTAGGCCCAGAACCAGCTCAGACTCCCCAACTTTGACAA 76
     |||:::|||  |||  ::|||:
 505 yArgArgAlaGlySerAlaArgAlaArgGlyProGluAlaAlaProA 522

  75 GAAGTGATCAGCTGAATCTTCAACCCATAGCAGAGCAAGAGCCCCCTGG 26
     ::::|:: |||::|:
 522 laArgGlnProGlyArgArgAlaArgLeuProAArgThrArgGlyProPro 538

  25 CTGCCCAGAAATTCAGCACTCTGG 2
     ||| ||| |||
 539 ArgProAlaProSerGluProTrp 546
```

```
XX
DR    WPI; 1999-069812/06.
DR    N-PSDB: AAV82909.
XX
PT    DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT    assembly of functional glutamate receptor subunits
XX
PS    Example 3; Columns 253-262; 203pp; English.
XX
CC    The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC    receptor subunit (NMDAR). The nucleic acid sequence does not contain
CC    the 366 5'-most nucleotides, by the insertion of 11 nucleotides between
CC    nucleotides 1300 and 1301, nor the 15 nucleotides at positions
CC    1960-1974, nor the 1061 3' nucleotides, as set forth in AAV82889. The
CC    cDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute
CC    to the formation of NMDA-activated cation-selective ion channels. In
CC    addition to being useful for the production of NMDA receptor subunit
CC    proteins, the nucleic acids are also useful as probes to identify and
CC    isolate nucleic acids encoding related receptor subunits. Functional
CC    glutamate receptors can be assembled from several NMDA receptor subunit
CC    proteins of one type (homomeric) or from combinations of subunit proteins
CC    of different types (heteromeric). The present invention also comprises
CC    methods for using such receptor subunits to identify and characterise
CC    compounds which affect the function of such receptors, e.g. agonists,
CC    antagonists and modulators of glutamate receptor function. The invention
CC    also comprises methods for determining whether unknown protein(s) are
CC    functional as NMDA receptor subunits.
XX
SQ    Sequence  1212 AA;

alignment_scores:
      Quality:  217.00              Length:       528
      Ratio:    0.995               Gaps:          36
Percent Similarity: 41.288          Percent Identity: 28.220

alignment_block:
US-09-786-136-4/rev x AAW87503  ..

Align seg 1/1 to: AAW87503 from: 1 to: 1212
```

```
1190 GCAGGTTTTGGTCCCAGTGGCGAGTAGCAGGAGTAGCAGGAGAGCCCTCGGCGCTGCG 1141
     |||:::                                              |||
791  AlaSerAlaGlyCysArgAlaSerProAlaHisArg...............       802

1140 GCCTGAGGGCGGCCAGGTCAA..GGCCGTCAGGCAAGTGCTGGGCCCCG 1094
     |||||||||:::       ||||||||||||
803  ......GlyArgProAlaArgThrSerArgProAlaArgProArgProA  817

1093 CAGCCAGC.......AATCAGCT.....TGGCCGCTGTCAGTTCCC 1059
     |::||:|||   :::|||||||           ||||||
817  laCysSerArgPheCysArgGlnProAlaThrTrpPro..........   829

1058 ACTCGTCGAAGGCAGCAGCGGCGCCGCCATCGGTCCTCCAGCCAGGCGAGTGCTACG 1009
                                          ||||||
830  ................................ArgArgAlaAlaAlaTh  834

1008 GCAGTGG..............CCCAGGTCGCCCGGCCCCAGGCCAGGTCGGT 977
                          |||:|:   |||   |||:::
834  rProTrpThrAlaProLeuAlaProSerArgIleGlyValAlaAlaAlaV  851

976  GCGGCCCAGCCCTTCCGCCCCATCGGAGGCCCGTCGTGTCGAGCCCCGC 927
     ||||||||     ||||||   |||::||||||
851  alAlaArgProHisArgProPro...AlaArgProArgGlyLeuAlaProAla 866

926  CAC...TGTCCACCTGGCCTGGCCCTCAGAGCCCGAGCTGGGGACGGG 880
     |||   |||||||||||||      |||||||||
867  HisAlaCysProProProThrArgProGlnSer..............    877

879  CTGCAGGAGGCCGTGCACTGTCACCCTGGCCAACACCAGGGCCCAGGAGAAG 830
     ...................................................
877  .................................................. 877

829  TGCCCAGGGCAATGAGGCCGAGGTGGGCGTGAGCGTTGGGCGTGCACGG 780
     |||||       ||||||      ||      :::|||
878  ....ArgAla.....ProArgAlaGlyAspArgGlnThrGlyValAla   890

779  CAAAGGGCGAGGCACGGACAGAGAGCGCTCCTGCGAGATGCCACAGCGGCG 730
     ::::|||           |||       ||       |
891  ArgArgLeuCysAlaGlyLeuArgSerProArg.....AlaAlaProAr  905

729  CAGAAGGGCGGCGTCCAGGGCGAAGGAGCCTGGTGCTCCTGCCAGCCGCAC 680
     ||||||||:::|||||||         |||    :::
905  gArgArgGlyArgPro...........CysProThrSerProGluC      917

679  CAAGGGCAGTAGTCGCTGCCTCACTGTTGTTGGCCTGGCCCCG 630
     |||||:::::        ||||||    |||||
917  ysArgValAlaGlnProGlyArgArgGlyArgGlyArgCysGlyProGlyThr 933

629  GGTCAGGGCTACAGAGTTTGTGACGGCGGCCCATGCTGAGCCGACAGGGA 580
     :::      |||       ||||||||||
934  AlaGlyGlyThrSerArgProProSerGlyProCys.ArgProArgAlaV 950

579  GGAGGGGGAGG............TGGGCCGGCGAGGACCTTC 551
     |||               |||
950  alThrThrAlaProPheLeuGluProThrAspProAlaAlaProSerSer 966

550  CGTGGGGAGTGGCAGGGTCCCCAAAGCACGCTTGAGTTGCCATTTTGT 501
     |||     |||:::|||  :::::|||       ||
967  Arg...SerSerArgSerProArgSerTrpArgThrCysArgCysSerVa 982

500  CCGAGTCCCAGGCGGCCTGCTGGGAGAGCCTCTCTGAAGTCCTTTAGAGTAG 451
     ||||||||::::       ||||||:::||:||
982  lArg.SerSerTrpProGlyLyArgProCys...............     992

450  ACAACAGGTGGGCAGGAGGAGCTGCATCTAGA............       421
     |||        :::||||||        :::|||
993  ThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCysProAL 1009

420  ....TGGCTGTGAGAGGCAGGAGGGGTTCTGGGGCTCAGCGGCTG 375
     |||              |||   |||
1009 aProTrp...........ProArgProSerLeuGlyPro.          1018

374  AGCTTGCACACCTGCAGGCGCCCCAGGCAGGACCTCCCTAGTAGTAGCATCC 325
```

```
1487 GGACTGTGGCCTCTCACCCTTCCACTG...........TGTTGCAG 1453
     ||||||  ||  ||||||||
677  GlyLeuAlaArgSerLeuLeuProLeuAlaThrAlaSerProCysArgAr 693

1452 GGAGCTCAGAGGA...........TGGAGGGGAGTGT...GAGTGGGGT 1418
     |||||||            |||||||:::   ||||||
693  gThrProThrGlySerGlyProThrArg.SerCysSerSerTrpGly    709

1417 GGGGAGGAGGGAGGAGCACCAGGCAGGCACTTTTGTGGGGGCAAGAGG 1368
     :::|||   |||||:::        ||||||:::
710  ThrGluArgHisArgAsnTrpArgGlnCys................Gl  720

767  GCTCCAGGGAGGGAGTAAGAAGCAG....................... 1343
     |   ||||||||||:::::        ::
720  ySerGlnGlySerAlaArgMetArgArgThrArgAlaAlaSerTrpThrS 737

1342 ..........................TGGGGGACAGGCTGG 1328
     |||   |||||||||
737  erThrThrTrpGluAlaSerThrCysCysTrpTrpProTrpGlyTrp    753

1327 CACCAGCCTACACTTTCCCAGCACCAAGAGGCCAGCCCTGGGCGCAGTG 1278
     |||         ||||||||||
754  ......ProCysTrpSer............SerProGlySerThr..    762

1277 ACTTGAGTGTGTTGGGGTGGGGCCAGCCAAGCAGCAGGGGCAGCCCTT 1228
     :::|||      |||||:::   |||||||
763  ..........TrpSerThrGlySerCysAlaThrArgCysSerProT    774

1227 CACACAGTTCCCTGGCCTGT...........AGCCACACCAGGTGTAGCT 1191
     :::|||      |||:::::||      :::||||||
774  hrHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 790
```

```
CC    also comprises methods for using such receptor subunits to identify and
CC    characterise compounds which affect the function of such receptors, e.g.
CC    agonists, antagonists and modulators of glutamate receptor function. The
CC    invention also comprises methods for determining whether unknown
CC    protein(s) are functional as NMDA receptor subunits.
XX
SQ    Sequence  1061 AA;

alignment_scores:
                     Quality:  217.00          Length:  528
                     Ratio:    0.995           Gaps:    36
     Percent Similarity: 41.288     Percent Identity: 28.220

alignment_block:
US-09-786-136-4/rev x AAW87504

Align seg 1/1 to: AAW87504    from: 1   to: 1061

1487 GGACTGTGGCTCGTCTCACCCTTCCACTG............TGTTGCAG 1453
     |||||| ||  |||||||||||| ||  |||                |||  |||
 526 GlyLeuAlaArgSerLeuProLeuAlaThrAlaSerProCysArgAr 542

1452 GGAGCTCAGAGGA............TGGAGGGAGTGT...GAGTGGGGT 1418
     |||          |||          |||||| :::||||  ::||||||
 542 gThrProThrGlySerGlyProThrTrpArg.SerCysSerSerTrpGly 558

1417 GGGAGGAGGAGAGAGCACAGCAGCCACTTTTTGTGGGGGCAAGAGG 1368
     |||:||::::  |||||::   ||||                     ||
 559 ThrGluArgHisArgAsnTrpArgGlnCys................Gl 569

1367 GCTCCAGGGAGGGAGTAAGAAGCAG................ 1343
     | |||||:::|:::::  |||::                  ||
 569 ySerGlnGlySerAlaArgMetArgArgThrArgAlaAlaSerTrpThrs 586

1342 ...................................TGGGGACAGGCTGG 1328
                                        |||  |||||||
 586 erThrThrTrpGluAlaSerSerThrCysTrpTrpProTrpGlyTrp 602

1327 CACCAGCTACACTTTCCAGCACCAAGAGGGCCAGCCCTGGGCGGCAGTG 1278
     | |||:::                              |||||||||
 603 ......ProCysTrpSer............SerProGlySerThr.. 611

1277 ACTTGAGTGTGTTGGGTGGGGGCCCAGCCCAAGCAGCAGGGGGCAGCCTT 1228
     |||:::::::::             :::        |||   |||||||
 612 ............TrpSerThrGlySerCysAlaThrArgCysProT 623

1227 CACACGTTCGCTGGGCTGT................AGCACAGCAGTGTAGCT 1191
     ::|||  |||:::                   |::  |||::::::|||
 623 hrHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 639

1190 GCAGGTTTTGGTCCCAGTGGCCAGTAGCAGGAAGAGCCCTCGGGCTGCG 1141
     ::|||   :::  |||||:::::::::||                |||
 640 AlaSerAlaGlyCysArgAlaSerProAlaHisArg................ 651

1140 GCCTTGAGGCGGCCGCCAGGTCAA...GGCCGTCAGGCAAGTGGTGGGCCCG 1094
     |||||||||:::||   :::||| |||||||   ::::||      |||
 652 ......GlyArgProAlaArgThrSerArgProAlaArgProAlaArgProA 666

1093 CAGCCAGC............AATCAGCCT...TGGCCCGTCTGTCAGTTCCC 1059
     ||::::|||    ::|||||   :::||||                |||||
 666 laCysSerArgPheCysArgGlnProAlaThrTrpPro.......... 678

1058 ACTCGTCGAAGGCAGCCGGCGCATCGGTGCTCCAGCCAGGCGAGTGCTACG 1009
                                        ||||||||
 679 ................ArgArgAlaAlaTh 683

1008 GCAGTGG................CCCAGGTCGGCCGCCCCAGGTCGGT 977
     |||                    |||   |||:::
 683 rProTrpThrAlaProLeuAlaProSerArgIleGlyValAlaAlaIlav 700

 976 GCGCCCCAGCCCTTCCGCCCCATCGGAGCCTCGGTGTCTGAGCCCCGCC 927
     ||||||       |||||       ||:::||||   ||:::||||||||
```

```
1019 :::          |||      |||    |||  ||:::   :::
     AlaArgCysProLeuGlyAlaProAla...Pro.ProAlaProAlaProT 1034

 324 ACAGCTACAGGCAAGTGAAGCAGGAGGGGCAGAGCTGACCTTACTTGT 275
     ||||       || :::   |||    ||:::||     :::|||
1034 hrAlaThrArgProAlaGlyAlaTrpArgArgArgSerArgCysAlaCys 1050

 274 CTGAAGGGCTCGAAGCCAGCA.................CCTCC 249
     ||  :::  |||:::|||||||                |||||
1051 ArgSerThrGlyArgProAlaArgArgAlaSerArgGlnGlyProProPr 1067

 248 GGGCATGGCCTGTGGAACCAAGGAGCAGTCTGTGGAGCCCCCACCACCA 199
     |||  ::|||   |||:::  |||  |||||:::  ||  |||   ||||
1067 oGly........SerThrAspSerAlaCysThrProThrProT 1081

 198 CCCTCTAC.........AGAGGGCTC.................ATAGAAAAAATTGCA 167
     ||   |||         ||:::|||                 |||
1081 hrCysHisCysAlaGlyGlyLeuSerValLeuThrPheHisProValThr 1097

 166 GTGTTCTCCCCAGCTTCTCCCTCT 143
     :::  |||::::|||:::||||||
1098 AlaThrAlaProGlySerProAla 1105
```

```
seq_name: /SIDS1/gcgdata/geneseqp-embl/AA1999.DAT:AAW87504

seq_documentation_block:
ID   AAW87504 standard; Protein; 1061 AA.
XX
AC   AAW87504;
XX
DT   23-FEB-1999 (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW   NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS   Homo sapiens.
XX
XX
PN   US5849895-A.
XX
PD   15-DEC-1998.
XX
PF   20-APR-1994;   94US-0231193.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI   Daggett LP, Lu C;
XX
DR   WPI; 1999-069812/06.
DR   N-PSDB; AAV82910.
XX
PT   DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT   assembly of functional glutamate receptor subunits
XX
PS   Example 3; Columns 261-268; 203pp; English.
XX
CC   The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC   receptor subunit (NMDAR). The nucleic acid sequence does not contain the
CC   860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891)
CC   between nucleotides 1300 and 1301, an additional 24 nucleotides
CC   (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in
CC   AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR
CC   subunits contribute to the formation of NMDA-activated cation-selective
CC   ion channels. In addition to being useful for the production of NMDA
CC   receptor subunit proteins, the nucleic acids are also useful as probes to
CC   identify and isolate nucleic acids encoding related receptor subunits.
CC   Functional glutamate receptors can be assembled from several NMDA
CC   receptor subunit proteins of one type (homomeric) or from combinations of
CC   subunit proteins of different types (heteromeric). The present invention
```

```
700 alArgProHisArgProPro...AlaArgProArgGlyLeuAlaProAla 715
926 CAC...TGTCCACCTGGCCTGGCCCTCAGAGCCCGAGCTGGGGACGGG 880
    |||  ||||||||||         |||||||||
716 HisAlaCysProProProThrArgProGlnSer................ 726

879 CTGCAGGAGGCCGTGGCACTGTCACCCTGGCCAACACCAGGGCCCAGAAG 830
726 .................................................. 726

829 TGCCCAGGCAATGAGGCCCAGGTGGGCTGAGGCTGGCCGCGTGCACCG 780
    |||||   ||||||  |||  |||||:::::
727 ....ArgAla......ProArgAlaGlyAspArgGlnThrGlyValAla 739

779 CAAAGGGCGGAGGCACGGCAGCAGGCCCTCCTGCGAGATGCGCACAGCGGCG 730
    :::||||    |||  :::|||      ||  |||
740 ArgArgLeuCysAlaGlyLeuArgSerProArg.....AlaAlaProAr 754

729 CAGAAGGGGGCGGCCGTCCAGGCGGAAGGAGCCTGGTGCCTCCTGCAGCCGCAC 680
    ||||||::::|||||                |||  :::
754 gArgArgGlyArgPro.............CysProThrSerProGluC 766

679 CAAGGGCAGGTAGTCATGGTCGCTCGCCTTCACTGTTGTTGGCCTGGCCCG 630
    |||||||::::|  |||||     |||  |||||
766 ysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGlyProGlyThr 782

629 GGTCAGGGTACAGAGTTTGTGACGGCGGCCATGCTGAGCCGACAGGGA 580
    :::  |||  :::||||||||| ::::|||      :::
783 AlaGlyGlyThrSerArgProProSerGlyProCys.ArgProArgAlaV 799

579 GGACGGGGAGG................TGGGCGGCGAGGACCTTC 551
    ||                        ||||||  :::
799 alThrThrAlaProPheLeuGluProThrAspProAlaAlaProSerSer 815

550 CGTGGGAGTGCAGGGTCCCCCAAAGCACCGTTGGAGTTGCCATTTTGGT 501
    |||  ||::::|||  |||::::|||             ::||
816 Arg...SerSerArgSerProArgThrArgThrCysArgCysSerVa 831

500 CCGAGTCCCAGGCGCGCTGCTGGAGAGCCTCTCGAAGTCCTTTAGAGTAG 451
    ||||  ||::: |||::|||:::||||
831 lArg.SerSerTrpProGlyGlyArgProCys................. 841

450 ACAACAGTGGGCAGGAGGAGCTGCATCTAGA................... 421
    ||  ||||||  |||  :::||
842 ThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCysProAl 858

420 .....TGGCTGTGAGAGGCCAGGAGGGGGTTCCTGGGGGGCTCAGCGGGCTCTG 375
    |||  |||  |||
858 aProTrp.................ProArgProSerLeuGlyPro. 867

374 AGCTGCACACCTGCAGGGCCCCAGGCAGGACCTCCCTAGTAGCATCC 325
    :::  ||||  |||  |||  |||::: :::
868 AlaAlaArgCysProLeuGlyAlaAlaProAla...Pro.ProAlaProAlaProT 883

324 ACAGTACAGGGCAAGTGAAGCAGGAGGGGGCAGAGCTGACCTTACTTGT 275
    ::::|||  :::  :::||||||||:::||||
883 hrAlaThrArgProAlaGlyAlaTrpArgArgArgSerArgArgCysAlaCys 899

274 CTGAAGGGCTCGAAGCCAGCA....................CCTCC 249
    :::  :::::||||||                         |||||
900 ArgSerThrGlyArgProAlaArgAlaSerArgArgGlnGlyProProPr 916

248 GGGCATGGCCGTGTGGAACCAAGGAGCAGTCTGTGAGCCCCACCACCCA 199
    ||||   ::: :::  :::|||      |||  |||
916 oGly........SerThrAspSerThrSerAlaCysThrProThrProT 930

198 CCCTCTAC......AGAGGGCTC............ATAGAAAAATTGCA 167
    ||  :::        ||||||            :::::  ::::::
930 hrCysHisCysAlaGlyGlyLeuSerValLeuThrPheHisProValThr 946

166 GTGTTCTCCCCAGCTTCTCCCTCT 143
    :::  :::|||:::|||||||:::
947 AlaThrAlaProGlySerProAla 954
```

OM of: US-09-786-136-4 to: PIR_71:*   out_format : pfs

Date: Sep 19, 2002 5:43 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09786136/runat_16092002_141243_19196/app_query.fasta_1.1592
-DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -LOOPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-ST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-GN=10 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-LEN=0 -MAXLEN=2000000000 -USER=US09786136_@CGN1_1_55
-CPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-136-4
Query length: 1511
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 64.780000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir1:QQBE3	-	218.50	231.98	1.2e-05	660	! BHLF1 protein - human herpesvir
pir2:T38236	-	218.00	225.37	1.1e-05	1611	! hypothetical protein SPAC23A1.
pir2:S06733	+	214.00	227.63	2.2e-05	620	! hydroxyproline-rich glycoprotei
pir2:T43556	+	210.50	224.44	3.6e-05	574	! Wiskott-Aldrich syndrome protei
pir2:T38819	+	210.50	224.44	3.6e-05	574	! wiskott-aldrich syndrome protei
pir2:T18535	+	210.50	219.70	3.3e-05	1151	! high molecular mass nuclear an
pir2:I78557	+	210.50	218.75	3.2e-05	1323	! N-methyl-D-aspartate receptor
pir1:C45219	-	210.50	218.58	3.2e-05	1356	! N-methyl-D-aspartate receptor
pir2:S49915	-	209.50	218.42	3.8e-05	1188	! extensin-like protein - maize
pir2:S27224	-	205.50	213.44	6.4e-05	1323	! N-methyl-D-aspartate receptor
pir1:QQBE31	+	205.00	217.84	7.5e-05	641	! BPLF1 protein - human herpesvi
pir1:QQBE8	+	199.50	201.15	0.0001	3149	! hypothetical protein - Deinococ
pir2:F75518	+	197.50	208.04	0.0002	839	! hypothetical protein - human h
r1:QQBE3	-	196.50	208.61	0.0002	660	! BHLF1 protein - human herpesvi
:A43932	+	196.50	198.25	0.0002	3020	! mucin 2 precursor, intestinal
:A41182	-	192.50	199.15	0.0004	1419	! collagen alpha 1(II) chain pre
:B41182	-	192.50	198.83	0.0004	1487	! collagen alpha 1(II) chain pre
pir2:T48814	+	192.50	196.98	0.0004	1952	! hypothetical protein 15R6.220
pir1:B45344	-	192.00	197.26	0.0004	1733	! probable nuclear antigen - sui
pir2:A54849	-	189.50	190.99	0.0005	2944	! collagen alpha 1(VII) chain pr
pir2:G86292	+	189.00	197.77	0.0006	1006	! hypothetical protein F7H2.17
pir1:CGHU6C	+	188.50	194.58	0.0006	1487	! collagen alpha 1(II) chain pre
pir2:A28996	-	188.50	204.05	0.0009	317	! proline-rich protein M14 precur
pir1:B40333	+	187.00	192.99	0.0008	1486	! collagen alpha 1(II) chain pre
pir2:B40505	+	187.00	191.11	0.0008	1958	! hypothetical protein - suid he
pir2:S22456	+	186.00	201.78	0.0011	350	! hydroxyproline-rich glycoprotei
pir2:A35363	-	186.00	197.01	0.0010	705	! synapsin I splice form a - huma
pir2:T04859	+	185.50	195.29	0.0010	839	! extensin homolog F28A21.80 - A
pir2:T45467	-	184.50	190.66	0.0011	1418	! collagen alpha 1(II) chain pre
pir2:E30411	+	182.50	193.28	0.0016	706	! synapsin Ia - bovine
pir2:JQ04405	-	182.50	190.22	0.0015	1106	! hypothetical 119.5K protein (u
pir2:A40333	-	182.50	188.18	0.0014	1492	! collagen alpha 1'(II) chain pr
pir2:S32975	+	182.00	196.92	0.0018	383	! gene BCRF2 protein - human herp
pir1:S29126	+	182.00	195.54	0.0018	469	! properdin precursor [validated
pir2:A59266	+	182.00	181.79	0.0014	3530	! unconventional myosin-15 - hum
pir2:C43674	+	181.00	191.76	0.0019	699	! US4 protein - human herpesvirus
pir2:A30411	+	180.50	191.18	0.0021	704	! synapsin Ia - rat
pir2:T46289	+	180.50	189.80	0.0020	862	! hypothetical protein DKF2p434A1
pir2:S28774	-	180.00	188.07	0.0021	1027	! collagen alpha chain - tube wo
pir2:T31611	+	180.00	185.12	0.0020	1585	! hypothetical protein Y50E8A.g
pir2:T42761	+	180.00	178.92	0.0018	3938	! Bassoon protein - rat
pir2:JC4163	-	179.00	188.50	0.0025	825	! DNA-binding protein 5E5 - ra
pir2:S45306	+	179.00	181.47	0.0022	2318	! notch 3 protein - mouse
pir2:T42567	+	178.00	177.53	0.0024	3534	! tegument protein 24 - equin
pir2:T15142	+	177.50	191.54	0.0033	418	! hypothetical protein T28F2.6

seq_name: pir1:QQBE3

seq_documentation_block:
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C;Accession: A03742
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
A;Reference number: A93065; MUID:85035713
A;Accession: A03742
A;Molecule type: DNA
A;Residues: 1-660 <BAN>
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667
A;Contents: annotation; protein coding region
C;Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399
C;Superfamily: human herpesvirus 4 BHLF1 protein

alignment_scores:
      Quality: 218.50   Length: 586
      Ratio: 0.967   Gaps: 34
Percent Similarity: 38.567   Percent Identity: 26.792

alignment_block:
US-09-786-136-4/rev x QQBE3

Align seg 1/1 to: QQBE3 from: 1 to: 660

```
1481 TGGCTCTCACCCTTCCACTGTGTTGCAGGAGCTCAGAGGATGGAGGGG 1432
     ||| :::     ||| ||  ||    || |||  |        ||||||||
 148 TrpSerAlaArgAsnProGlyCysProArgThr.............. 158

1431 AGTGTGAGTGGGTGGGAGGAGGAGGAGGAGACACAG............. 1395
     |||:::||:||||:|    :|||||||       ||||||
 159 ...........TrpArgArgSerGlyAlaGlnArgGlyHisProP 171

1394 ...........GCAGGCACTTTTGTGGGGCAAGAGG 1368
               ::||||             |||
 171 roProGlyAlaGlyGlnArgProSerGlyProThrGlyArgProAla 187

1367 GCTCCAGGGAGGGAGGAGTAAGAAGCAGTGGGGAGCAGAGGCTGGCACCA... 1323
     ||| ||          |||  ||||||
 188 AlaProGly..........AlaProGlyThrProAlaAlaProGlyPr 200

1323 .......................... 1323
 200 oGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgG 217

1322 ..........GCCTACACTTCCCAGCACCAAGA 1299
               |||    :::|||        |||
 217 lySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArg 233

1298 GGCCAGCCCTGG..........GCGGCAGTGACTTGAGTGTGTTGGGT 1261
     ||| |||             ::|||        :::||
 234 GlnGluProArgLeuProGlnAspProLeuAlaAlaAlaGlnArgCysProAl 250

1260 GGGGGCCGCCCCAGCCAGCAGGGGCAGCC 1230
     |||  :::||||::::|||||||||||
 250 aGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgA 267

1229 ...........TTCACGTTCGCTGGCGTGT.......... 1209
                 :::||::::          ||||||
```

```
267 rgrProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 283

1208 ......AGCACAGCAGTGTAGCTGCAGGTTTGCCCAGTGGCGCAGT 1165
           ::!!!!::    !!!    !!!!!
284 TrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAla.. 299

1164 AGCAGGAAGAGCCCTCGGGTCGCGGCCTTGAGGGGCGGCCAGGTCAAGGCC 1115
         !!! !!!!!!  !!!   !!!    !!!!
300 .........GlyGlnArgProSerGly...ProThrGlyGlyA 310

1114 GTCAGGCCAAGTCTGGG........CCCGCAGCCAGCAATCAGCCT 1077
     !! !!!!::  !!     !!!::    !!!!!
310 rgrProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326

1076 TGGCCCGTCGTCAGTTCCCACTCGTCGAAGGCAGCGGCCATCGGTGCT.. 1029
     ::!!!!!!!!!::
327 GlyGlyAlaAlaValProSer.................GlyAlaTh 336

1028 ......CCAGCCAGGCGAGTGCTACGGCAGTGGCCCAGTGCCGGCCCCG 986
          !!!   !!!           !!!!!!!!!
336 rProHisProGluArg................GlySerGlyProA 346

985 CAGGTCGGTGGCGCCCAGCC..CTTCCGCCCATCGGAGG........... 948
     !!!!!::::  !!!!!!!!!:::!!::
346 laAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArg 362

947 ..CCTCGGTGTCTGAGCCCCGCC..CACTGTCCACCTGCCTGCCCCC 904
     !!!::    !!!:::  !!!!!! !!!    !
363 LeuProGlnAspLeuAlaLeuAlaAlaGlnArgCysProAlaGly......Pr 377

903 TCAGAGCCCCGAGCTGGGGACGGGCTGC...........AGGAGGCCGT 866
        !!::!!!!  :::!! !!!  !!    !!!!!!!!!
377 oProProThrArgSerGlyAlaAlaAlaGlnAlaArgThrHisArgArgProP 394

865 GGCACTGTCACCCTGGCCACCACCAGGGCCCAGAAGTGCCCAGGCAATG 816
        !!!  !!    !! ::  ::!!!!!!  !!!!!!!!!
394 roGlyCys...ProArgSerAlaArgAsnProGlyCysProArgThr.... 408

815 AGGCCGAGGTGGGGCTGAGGCTGGCGCGGGTCACGCAAAGGGCGAGGCA 766
                            !!! !!!!!!!:::
409 ..............................Trp.........ArgArgSerGlyAl 415

765 CGGCAGAGGCGCTCGTGCAGAGATGCCACAGCGGCGGCCAGAAGGGGCGCGTC 716
     !!!!!!  !!!!!!  !!!!!!!!    !!!
415 aGlnArgGlyHisProProProGlyAlaGlyAlaGlnArgProSerGlyArgProT 432

715 CAGGGCGAAGAGACCTGGTGCCTCCTGCAGCCGCACCAAGGGCAGGTAGT 666
        !!!!!!!!!  :::!! !!!  :::  !!!!!!
432 hrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla...... 447

665 CATGGTGCTGCCTTCCACTGTGTTGCCTGCCCCGGTCAGGGCTACAG 616
                 !!!!!!!!!!!!!  !!!
448 ...............ProGlyProGlyGlyAlaAl 455

615 AGTTTGTGACGGCGGCCCATGCTGAGCCCGACAGGAGGACGGGGAGGTG. 567
       !!!   :::!!!!::        :::!!!   ::!!
455 aValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaA 472

566 ........GGCGGCGAGGAGACCTTCCGTGGGAGTGGCAGGGTCCC...... 531
          :::!!!:    !!!!!!  !!!  !!!   !!!
472 spProProAlaAlaAlaArgLeuProProGluProProGlnGluProArgLeu 488

530 CCAAAGCAC......GCTTGGAGTTGCCATTTGTCCGAGTCCCAG 490
     !!!::        !!!    !!!  !!!!!  !!!
489 ProGlnAspLeuAlaLeuAlaAlaGlnArgCysProAlaGlyProProProTh 505

489 GCGGCCTGCTGGAGAGCCTCTCTGAAGTCCTTTAGACTGAGACAACAGTGGG 440
        :::
505 rArgSerGlyAlaAla.............................A 511

439 CAGAGGAGCTGCATCTA......GATGGCTGTGAGAGGCAGGAGGGGTTC 396
     !!!!!!! !!!    !!!!!    ::!!!
439 laGlnArgThrHisArgArgProGlyCysProArgSerAlaArgAsn 527
```

```
395 CTGGGGGCTCAGCGGCGGCTCTGAGTCGTCACACCTGCAGGGCCCAGGCAGG 346
    !!!::: !!!::   !!!     ::!!::   !!!!!!
528 ProGlyCysProArgThrrpArgArgSerGlyAlaGlnArgGlyH1 544

345 ACCTCCCTAGTAGTAGCATCCACAGCTACAGGGCAGTGAAGCAGCAGGAGGG 296
    !!! !!!  :::!!::    !!!    !!!
544 sPro.ProProGlyAla.......GlyGlnArgProSerGlyProThrGlyGly 558

295 GGCAGAGTCGACCTTACTTGTCTGAAGGCTCGAAGCCAGCACACTCCGGG 246
    !!!!!!                  !!!!!    !!!!!
559 GlyArgProAlaAlaProGlyAlaProGlyThr...ProAlaAlaProG1 574

245 CATGGCCGTGTGGAACCAAGGAGCAGTCTGTGGAGCCCCACCACCCACCC 196
    !   :::   !!!:::        !!!    !!!    !!!
574 yProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 591

195 TCTACAGAGGGCTCATAGAAAAAAATTGCAGTGTGTTCTCCCAGCTTCTCC 146
                            !!!!!!:::!!!!!
591 rgGlySerGly..................ProAlaAspPro 598

145 TCTATCGACTTAACAGTAGATTGCCTTGCTAGGCCCAGAACCAGCTC 96
                              :::!!!!!!!
599 ProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGl 615

95 AGAC..................................TCCCAACTTTGACAAGAA 73
   :!!!                                  !!!!!!!
615 nAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgS 632

72 GTGATCAGGCTGAATCTTCAACCCATAGCACAGCAAGAGCCCCTGGCTG 23
   !!! !!    !!!!!!      !!!!!!
632 erGlyAlaAlaAlaGlnArgThrHis.........ArgArg.ProProGlyCy 646

22 CCCA 19
    !!!!
646 sPro 647
```

seq_name: pir2:T38236

seq_documentation_block:
hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38236
R;Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: 221780
A;Accession: T38236
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1611 <MUR>
A;Cross-references: EMBL:AL021813; PIDN:CAA16991.1; GSPDB:GN00066; SPDB:SPAC23A1.17
A;Experimental source: strain 972h-; cosmid c23A1
C;Genetics:
A;Gene: SPDB:SPAC23A1.17
A;Map position: 1

alignment_scores:
             Quality: 218.00   Length: 460
               Ratio: 0.991   Gaps: 26
    Percent Similarity: 47.826   Percent Identity: 26.957

alignment_block:
US-09-786-136-4 x T38236

Align seg 1/1 to: T38236 from: 1 to: 1611

```
314 CCTGTAGCT..GTGGATGCTACTACTAGGGAGGTCCTGCCTGGGCCCT 360
    !!!!!::::    !!! :::   !!!!!::::!!!!!!
812 ProValSerIleValThrSerGlyGlyArgProAlaLeuProAspGluMe 828

361 GCAGGTGTGCAGCTCAGCAGCCCGCTGAGCCC...CCAGGAACCCCTCCTG 407
```

```
         ;     |||||::  ::::::||   |||::    |||||||
 828 tAlaSerProSerSerIleGlyHisProLeuProSerProProA 845

 408 CCTCT........................CACAGCCATCTA 424
     ||:::                         :: |||||
 845 laAspPheAsnSerLeuAsnValAspPheTyrProHisSerTyrLeu 861

 425 GATGCA...GCTCCTCTGCCC................................ 442
     :::::: ||||||| |||
 862 GluSerProAlaProGluProGlnProSerTyrGluGluSerPheAs 878

 443 ...ACTGTTGTCTACTCTAAAGGACTTCAGAGGAGGCTCTCCAGCAGGCG 488
     ||||||:::::::   :: ||| |||::::
 878 nAlaThrValIleHisAlaProThrProSer.ThrAlaThrPheGlnGly 894

 489 CCTGGAGCTCCGACCAAAATGGCAACTCCAAGCCTGCTTTGGGGACCCT 538
     :|||  :::::::::||||||||
 895 HisProThrIleSerAsnValAlaThrPro.......................... 904

 539 GCCACTCCCACGGAAGGTCTCTGCGCCACCTCCCGTCCTCCCGTGTG 588
     |||||| ::    ::    :: |||||
 905 .ProLeuLysGlnAspValThrGluSer......LysAlaSerProValA 919

 589 GCTCAGCATGGGCCGCGTCACAAACTCTGTAGCCGTGACCCGG...... 632
     :::::|||  ::::   :|||||:||||||||
 919 laAspAlaSerAlaThrHisGlnSerThrGlyLeuThrGlnGluIle 935

 633 ...GCCAGGCCAACAACAGTGAAGGCAGCAGCCATGACTACCTGCCCTTG 679
     :::::::::||::   ||| |||
 936 ThrGlnLeuGlySerAsnMetArgLeuProThrLysThrArgProSe 952

 680 GTGCGGCTGCAGGAGGCACCAGCGCTCCTTCCGCCTGGACGCGCCCTCTG 729
     |||::  :::::|||                    |||||||
 952 rAsnAspGlyArgLysAlaSerGlyPro...........ArgProAlaA 965

 730 CGCGCCGTGCGCCATCTCGCAGGAGGCGCCTCTGCGCTGCCTCGCCCTTG 779
     ||||                    ::::::||   |||||
 965 laPro...........................ProSerIleProProProLeu 973

 780 CCG.............TGCACCGCGCAGCCTCAGCCCCACCTCGGCCTCA 817
     |||            ::::||||:|||  ||| ||::::|||
 974 ProValSerAsnIleLeuSerSerProThrSerProGluProLysAspHi 990

 818 TTGCCCTGGGCACTTCTGGGCCTGGTGTTGGCCAGGGTGACAGTGCCAC 867
     | ||:::        ||| |||
 990 sProProSerAla........ProLeuSerLysProValSerThrSerProA 1005

 868 GGCCTCCTGCAGCCCGT..CCCCCAGCTCGGGCTCTGAGGGGCCAGGCC 914
     ||| |||||  |||||::   |||  :::::||
 1005 laAlaProLeuAlaArgValProProValProLysLeuSerSerLysAla 1021

 915 AGGTGGACAGTGGGCGGGGCTCAGACACCGAGGCCTCCGATGGGCCGGAA 964
                              |||||||||
 1022 ..............................ProProVal.......... 1024

 965 GGGCTGGGCCGGCACCGACCTGCGGGGGCCGACCTGGCCCACTGCCGTAGC 1014
     |||::::::  ||| ||| |||:::|||
 1025 ......ProLeuProSerAlaAspAlaProProIleProValPro.... 1037

 1015 ACTCGCCTGGCTGGAGCACCCGATGCGCCGCTGCCTTCGACGAGTGGGAAC 1064
     ||::::|||           :::::::::
 1038 ..............SerThrAlaProProValProIleProThrSerThr 1049

 1065 TGACACGCGGCCAAGGCTGATTGCTGGCTGC.................... 1094
     ||:::                        |||:::
 1050 ProProValProLysSerSerGlyAlaProSerAlaProProProVa 1066

 1095 ...GGGCCCAGCACTTGCCTGACCGGCCCTGACCTGGCCGCCCCTCAAGGC 1140
     ::::|||||||  ::: ||||:: |||::  |||||||:
 1066 lProAlaProSerProSerGlyIleProSerIle...ProSerGly 1081

 1141 CGCAGCCCGAGGGCTCTTCCTACTGCGCCACTGGGACCAAAACCTGC 1190
     |||:::    ||| ||||||||||::
```

```
1082 AlaProProValProAlaProSerGlyIleProProValProLysProSe 1098

1191 A...GCTACACCTGCTGTGCTACAGCCCAGCGAACGTGAAGGCTGCCC 1237
     : ||:::||| ||| ::||||| :|||
1098 rValAlaAlaProProLysProSerValAlaValProProValP 1115

1238 CCTGCTGCTTGGGCTGGCGCCCCACCCACACACTCAAGTCACTGCC... 1284
     ||::|||      ::::::||:::::||||
1115 roAlaProSerGlyAlaProProValProLysProSerValAlaAlaPro 1131

1285 ..........GCCCAGGCTGGCCT.................CTTGG 1304
     :::|||     |||
1132 ProValProSerGlyAlaProProValProLysProSerValAl 1148

1305 TGCTGGGAAAGTGTAGGCTGCTGTGCCAGCCTGTCCCCACTGCTTCTTACT 1354
     :|||    ||| |||     :::::  ||| :::   |||||
1148 aAlaProProValProAlaProSerGlyAlaProProValProLysProS 1165

1355 CCCTC.................................CCTGGAGCCCTCTTGCCCCCCACAA 1383
     |||||                                |||::::|||||
1165 erValAlaAlaProProValProAlaProSerGlyIleProProVal 1181

1384 AAAGTGCCTGCCTGTGCTCTCTCCCTCTCCCACCCCACTCACACTCC 1433
     |||||   :::::||| :::::::|| |||::::
1182 ProLysProAla..AlaGlyValProProValProProProSerGluAla 1197

1434 CCT........................CCATCCTCTGAGCTCCTGCAACACAGTGGAAGGG 1471
     |||                        |||||| :::|||     ::: :::
1198 ProProValProLysProSerValGlyValProProValProProProSe 1214

1472 TAGAGAGCCACCAGTCCCCAAATCCT 1496
     || |||:::|||
1214 rThrAlaProProValProThrPro 1222

seq_name: pir2:S06733


seq_documentation_block:
hydroxyproline-rich glycoprotein precursor - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C;Accession: S06733
R;Keller, B.; Lamb, C.J.
Genes Dev. 3, 1639-1646, 1989
A;Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein ge
A;Reference number: S06733; MUID:9128263
A;Accession: S06733
A;Molecule type: DNA
A;Residues: 1-620 <KEL>
A;Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein


alignment_scores:
      Quality: 214.00          Length:      424
        Ratio: 1.251             Gaps:       22
Percent Similarity: 40.330    Percent Identity: 27.358

alignment_block:
US-09-786-136-4 x S06733

Align seg 1/1 to: S06733   from:  1 to: 620

269 CTTCAGACAAGTAAGGTCAGCTCTGCCCCTCCTTCACTGCCTGT 318
    | ||||||::::  ||| ||:::||::||||||||::
 13 LeuGlnSerThrAlaIleLeuSerLeu...................Va 22

319 AGCTGTGGATGCTACTACTAGG..............................G 341
    | |||||::||||||||||:::
 22 lAlaAlaGluAlaThrThrGlnTyrGlyTyrLeuProProProValT 39

342 AGGTCCTGCCTGGGCCCTGCCAGGTGCGCAGTCAGACCCCGCTGAGCCC 391
    :::::::                  |||::::   |||
```

```
  39 hrSerGlnProProProSerIleGlyLeuProProSerAlaPro  55      1221 ACGTGTGAAGGCTGCCCCTGCCTGTTGGGCTGGCGGCCACCCAACACA 1270
                                                                  ::: |||||||  :::::::  :::::::  |||||||: |||
 392 CCAGGAACCCCTCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCC 441    280 oThrThrTyrSerPro.ProProProThrTyrSerProProProSerPro 296
      ||||||||   ||| ||||:::|||:|   ||| ||
  56 ThrThrThrProPro...SerArgGlyHisVal.........ProSerPr 68     1271 CTCAAGTCACTCGCCGCCCAGGGCTGGCCTCTTGGTGCTGGGAAAGTGTAG 1320
                                                                  ::: |||  |||| ||||||  |||
 442 CACTGTGTCTACTCTAAAGGACTTCAGAGAGGCTTCCAGCAGGCGCCT 491    297 IleTyrSerProProProAla................... 304
      ::::::::  ::|||  |||||||   |||:||
  68 o.................ArgHisAlaProProArgHisAlaT 77         1321 GCTGGTGCCAGCCTGTCTCCCCACTGCTTCTTACTCCCTGCCTGAGCCCT 1370
                                                                  |||| ::::: ::|||:::|||::||
 492 GGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGGACCCTGCC 541    305 ....TyrSerProSerProProThrProThrProThrPheSerProP 320
      ::  :::::||:|| ::::|||||||  |||
  77 yrProProProSerHisGlyHisGlyHisLeuProProSerValGlyPro... 92    1371 CTTGCCCCCACAAAAAGTGCCTGCCTGCTCTCCTCCTCCTCCTCCTCCACC 1420
                                                                  ||| ::: |||||||||  |||||||
 542 ACTCCCACGGAAGGT........CCTCGCCGC........CCACCTCC 573    320 roProProAlaTyrSerProProProThrTyrSerProProProThr 336
      ||| :::|||  :::||| ||||||||   |||||||
  93 ProProHisArgGlyHisLeuProProSerArgGlyPheAsnProProPr 109    1421 CCACTCACACTCCCTCC 1438
                                                                  ||| |||||||||
 574 CCGTCCTCC.CTGTCGGCTCAGCATGGCCCGCCGTCACAAACTCTGTAGC 622    337 TyrLeuProLeuProSer 342
      |||  :::||| :::||| |||||  ||||||||
 109 oSerProValIleSerProSerHisProProSerTyrGlyAla...P 125
                                                             seq_name: pir2:T43556
 623 CCTGACCC........GGGCCAGGCCAA...CAACAGTGAAGGCAGCCACCA 663
      :  ||| ||||||||||:::  ::  |||||||||                 seq_documentation_block:
 125 roProProSerHisGlyProGlyHisLeuProSerHisGlyGlnArgPro 141    Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
                                                             C:Species: Schizosaccharomyces pombe
 664 TGACTACCTGCCCTTGGTGCGGCT.....GCAGGAGGCAC......... 698    C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
      ||:::  |||  ||| :::|||||                            C:Accession: T43556
 142 ProSerProSerHisGlyHisAlaProSerGlyGlyHisThrProPr 158    R:Zankel, T.C.; Ow, D.W.
                                                             submitted to the EMBL Data Library, December 1997
 699 .....CAGGCTCCTTCCGCCTGGACGGCCCCTTCTGCCGCCGCTGTGCG 741    A:Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pomb
      ::::|||   :::  :::|||||||   |||||||                 A:Reference number: Z22575
 158 oArgGlyGlnHisProProSerHisArgGlyProSerProPro...... 172    A:Accession: T43556
                                                             A:Status: preliminary; translated from GB/EMBL/DDBJ
 742 CATCTCGCAGGAGC..............GCCTCTGCCGTGCGCCT 770    A:Molecule type: DNA
      |||||||:::::::          |||                         A:Residues: 1-574 <ZAN>
 173 ..SerArgHisGlyHisProProProThrTyrAlaGlnProProPro 188    A:Cross-references: EMBL:AF038575; PIDN:AAB92587.1
                                                             A:Experimental source: strain JS21
 771 CGCCCTTTGCCGTGCACCGCGCCAGCCTCAGCCCCACCTCGGCCTCATTG 820    C:Genetics:
      |||:::  :::|||   ||||||   :::                       A:Gene: wsp1
 189 ThrProIleTyrSerProSerProGlnValGlnProProProThrTyrSe 205    A:Map position: I
                                                             A:Introns: 72/3; 519/3; 564/1
 821 CCCTGGGCACTTCTGGGCCCTGGTGTTGGCCAGGGTGACAGTGCCACGGC 870
      |||  |||                                            alignment_scores:
 205 rPro..............................ProProP 209          Quality:      210.50            Length:       452
                                                                 Ratio:       1.132            Gaps:          22
 871 CTCCTGCCAGCCCGTCCCCAGCTCGGGCTCTGAGGGGCCAGGCCAGGTGG 920        Percent Similarity:   41.150     Percent Identity:  25.664
      ||  :::::::|||
 209 roThrHisValGlnPro................... 214              alignment_block:
                                                             US-09-786-136-4 x T43556    ..
 921 ACAGTGGGCGGGGCTCAGACACACCGAGGCCTCCGATGGGGCCGAAGGGCTG 970
      |||||| |||||  |||                                    Align seg 1/1  to: T43556   from: 1  to: 574
 215 ..............ThrProSerProProSer.ArgGly......... 222
                                                             234 CCACAGGCCATGCCCGGAGGTGCTGGCTTCGAGCCCTTCAGACAAGTAAG 283
 971 GGCGGCACCGACCTGCGGGGCCGACCTGGGCCACCTGGCCTAGCACTCGC 1020        |||:|||   ||||||||   :::|||  |||
 223 ....HisGlnProGlnProThrHisArgHisAlaProThrHis 237      159 ProSerAlaProProValProGlyLysGluAsnTyrAsnAlaValGl 175
      |||:|| |||                                                   
1021 CTGGCTGGAGCACCGATGCGCGCTGCCTTCGACGAGTGGGAACTGACAG 1070    284 GTCAGGCTCGCCCCCCTCCGCTTCACTTGCCCTGAGCTGTGGATGCTAC 333
      :::  |||                                                 ||| |||   :::|||   ||||||
 238 ArgHisAlaProProThrHis.................... 244        175 yCerLysSerProAsnGluProGluLeu.......... 184
                                                             
1071 CGGCCAAGGCTGATTGCTGGCTGCGGGGCCCAGCCACTTGCCTGACGGCCTT 1120    334 TACTAGGGAGGTCCTGCCTGGGGCCTGCAGGTGTGCAGCTCAGAGCCCG 383
      ::|||                                                    ::||||
 245 ..............GlnProS 247                            184 ....................... 184
                                                             
1121 GACCTGGCGCCCCTCAAGGCCCAGAGCCCAGGCGCTCTTCCTGCTACTGCG 1170    384 CTGAGCCCCAGGAACCCCTCCTGCCTCTCCACAGCCATCTAGATGCAGCT 433
      ||| ||                                                   |||:|||   |||   :::|||   |||:|||
 247 erProLeuArgHisLeuProSerProArgArgGlnProGlnProPro 263      185 LeuAsnSerLeuAspProSerLeuIleAspSerLeuMetLysMetGlyIl 201
                                                             
1171 CCACTGGGACCAAAACCTGCAGCTACACCTGCTGTCTGCTACAGCCCAGCGA 1220    434 CCTCTCCCCACTGTTGCTCTACTCTAAAGGACTTCAGAGAGGCT...CTCC 480
      ::|||  |||   |||                                         ||||||  |||
 264 ThrTyrSerProProProAlaTyrAlaGlnProGlnSerProGlnProSer 280    201 eSerGlnAspGlnIleAlaGluAsnAlaAspPheValLysAlaTyrLeuA 218
```

```
1291 GGCTGGCCTCTTGGTGCTGTGGGAAAGTGTAGGCTGGTGCCAGCCTGTCCCC 1340
              |||
 442 yAla.......................................ProAlaAlaProP 448

1341 CACTGCTTCTTACTCCCTCCCTGGAGCCCTTGCCCCCCACAAAAAGTGC 1390
     |||||   |||       ::::||::|||:|||||
 448 roLeu.......Pro.ProSerAlaProIleAlaProProLeuProAl 461

1391 CTGCCTGCTGTCTCTCCTCTCCCACCCCACTCACACTCCCTCCAT 1440
     |  :  ::   |||  |||||||||  ::  ::||||||
 461 aGlyMetProAlaAlaProLeuProProAlaAlaProAlaProProP 478

1441 CCTCTGAGCTCCCTGCAACACAGTGGAAGGGTAGAGAGCCACAGTCCCCA 1490
              ||||||                    |||  ::||||
 478 roAlaProAlaProIa...........................ProAlaAlaPro 487
```

seq_name: pir2:T38819

alignment_scores:
             Quality: 210.50   Length: 449
             Ratio: 1.132   Gaps: 23
Percent Similarity: 41.425   Percent Identity: 27.171

alignment_block:
US-09-786-136-4 x T38819   ..

Align seg 1/1 to: T38819   from: 1   to: 574

```
 234 CCACAGGCCATGCCCGGAGGTGCTGGCTTCGAGCCCTTCAGACAAGTAAG 283
     |||:||:|||   ||||||||||  |:||||     :|:::|
 159 ProSerAlaAlaProProValProGlyLysGluAsnTyrAsnAlaValGl 175

 284 GTCAGCTGCGCCCCTCCTGCTTCACTTGCCCTGTGGGATGCTAC 333
     |||   |||      ||||||::::||:||||||
 175 ySerLysSerProAsnGluProGluLeu................... 184

 334 TACTAGGGAGGTCCTGCCTGGGGCCCTGCCAGGTGTGCAGCTCAGAGCCCG 383
     ................................................... 184

 384 CTGAGCCCCCAGGAGACCCCTCCTGCCTCTTCACAGCCATCTAGATGCAGCT 433
     :::|||::  |:|   |||  ||::::|||||
 185 LeuAsnSerLeuAspProSerLeuIleAspSerLeuMetLysMetGlyI11 201

 434 CCTCTGCCCACTGTGTCTACTCTCAAAGGACTTCAGAGAGGCT...CTCC 480
     ::   |||    ||||||||      ||||||       :::|||
 201 eSerGlnAspGlnIleAlaGluAsnAlaAspPheValLysAlaTyrLeuA 218

 481 AGCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGG 530
     ::::::|| ||||||| |||::|  |||||||| |||||
 218 snGluSerAlaGlyThrProThrSerAlaProIle........... 232

 531 GGGACCCTGCCACTCCCACGGAAGGTCCTGCCGGCCCACCTCCCCGTCCT 580
```
```
 481 AGCAGGCGGCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGG 530
     ::::::||||| |||::  |||||||| |||:|||
 218 snGluSerAlaGlyThrProThrSerAlaPro............. 230

 531 GGGACCCTGCCACTCCCACGGAAGGTCCTGCCGCCGCCCACCTCCCCGTCCT 580
     |||||||| ||:::||  |:::|||  |||::::
 231 ......ProIleProProSerIleProSerSerArgPro...... 242

 581 CCCTGTCGGCTCAGCATGGGCGGCGGCTCACAAACTCTGTAGCCCTGACCC 630
     :::||| |: :||  |||  ||||||||||||
 242 ...................................................... 242

 631 GGGCAGGCCAACACAGTGAAGGCAGGGACCCATGACTACCTGCCCCTGG 680
     |||         ::|||::||  ::   |||
 243 ..GluArgValProSerValSerAlaProAlaProProProIlePro... 257

 681 TGCGGCTGCAGGAGGCACCCAGGCTTCCTTCCGCCTGGACGCGCCCTTCTGC 730
              ||||||      |||       |||
 258 ........................ProProSerAsnGlyThrValSerSerPr 267

 731 GCCGTGTGCCATCTCGGAGGAGCGCCTTCGCCGTGCCTCGCCCTTG.. 779
     |||        :::::  |||||||||||:
 267 oPro..................AsnSerProProArgProIleA 276

 780 ..............CCGTCACGGCGCAGCCTCAGCCCACCACCTCGG 812
     |||::  ||||||
 276 laProValSerMetAsnProAlaIleAsnSerThrLysProProLeu 292

 813 CCTCATTGCCCTGGGCACTTCTGGGCCTGTGTTGGCCAGGTGACAGT 862
     |||    ::::::  |||  ::|||||   :::::
 293 ProProProSerSerArgValSerAlaAlaLeuAla...AlaAsnLy 308

 863 GCCAGGCCTTCCTGCAGCCCGTCCCCAGCTCGGGCTTGAGGGCCAGG 912
     ||||||||     |||||:::||||||
 308 sLysArgProProProProProSerArgArgAsnArgGlyLysP 325

 913 CCAGGTGGCAGTGGGCGGGCTCAGACA............. 941
     |||::::::::::::::
 325 roProIleGlyAsnGlySerSerAsnSerSerLeuProProProPro 341

 942 ...............CCGAGGGCTTCCGATGGGGCG 961
     |||  |||||||:||||
 342 ProProArgSerAsnAlaAlaGlySerIleProLeuProProGlnGlyAr 358

 962 GAAGGGC..........TGGGCGGCACCGACCTGCGGGGGCCG 993
     |:::||||
 358 gSerAlaProProProProProArgSerAlaProSerThrGlyArg 375

 994 GACCTGGGCCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGCCG 1043
     |||  |||||| |||:::|||       |||
 375 lnProProLeu.SerSerSerArgAlaValSerAsnPro........ 388

1044 CTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGGCTG 1093
     ................................................... 388

1094 CGGGGCCAGCACTTGCCTGACGGCCTTGACGGCGCCCTTCAAGGCCGC 1143
     |||||     ||| ::||  ||   |||||||
 389 ....ProAla..........ProProProAlaIlePro...GlyArg 398

1144 AGCCCGAGGGCTCTTCCTGCTACTGCGCCCACTGGA...CCAAAACCTGC 1190
     |||   ||||||||||          ::  :::
 399 SerAlaProAlaLeuPro.......ProLeuGlyAsnAlaSerArgTh 412

1191 AGTACACCTGCTGTGTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCT 1240
     :::|||||| |||  :::|||   :::      |||||||
 412 rSerThrProProVal.ProThrProProSer.......LeuProPro 425

1241 GCTGCTTGGGCTGGCGCCCCACCCAACACTCAAGTCACTGCCGCCCAG 1290
     :::|||   :::  |||||||||             |||||||
 426 SerAlaProProSerLeuProProSerAlaProProSerLeuProMetGl 442
```

```
233  .....ProPro.SerIleProSerSerArgProGluArgValProS  247
             |||  |||      |||     ||||||
581  CCCTGTCGGCTCAGCATGGGCCGCGTCACAAACTGTGTAGCCCTGACCC  630
     |||||||||                                 |||
247  erLeuSerAlaProAlaProProIle.................Pro  257

631  GGGCCAGGGCCAACAACAGTGAAGGCAGGCGACCATGACTACCTGCCCTTGG  680
     |||
258  ProPro.....................................  259

681  TGCGGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACG......CGCC  723
     :::::||||||             ||||||::::         ||||
260  .......SerAsnGlyThrValSerSerProProAsnSerProProArgP  274

724  CTTCTGGCGCCGCTGTGCCATCTCGAGGAGCCGCCTCTGCCGTGCCTCGC  773
     ||  |||||||          ::
274  roIleAlaProValSerMetAsn...................  281

774  CCTTTGCCGTCGCACGCGGCCAGCCTCAGCCCCACCCTCGGCCTCATTGCCC  823
     ||:::        :::||  |||||    |||
282  .....ProAlaIleAsnSerThrSerLysProProLeuProProProSe  296

824  TGGGCACTTCTGGGCCCTGGTGTTGGCCAGGTGACAGTGCCACGGCCTC  873
     :::::   |||  ::|||||||  :::::  ||||||
296  rSerArgValSerAlaAlaAlaLeuAla..AlaAsnLysLysArgProP  312

874  CTGCAGCCGCGTCCCCAGCTCGGGCTCTGAGGGGCCAGGCTGGACA  923
     ||    ||||||::::::
312  roProProProProSerArgArgAsnArgGlyLysProProIleGly  328

924  GTGGCGGGGCTTCAGACA.......................  941
     ||::::||||:
329  AsnGlySerAsnSerLeuProProProProProProArgSe  345

942  ...........CCGAGGCCTCCGATGGGGCGGAAGGGC.....  968
             |||  ||||||::|||||||::::::
345  rAsnAlaAlaGlySerIleProLeuProProGlnGlyArgSerAlaProP  362

969  .......TGGGCGGCACCGACCTGCCGGGCGGACCTGGGCCA  1004
             ::|||||||:::  |||  |||  |||
362  roProProProArgSerAlaProSerThrGlyArgGlnProProPro  378

1005 CTGCCGTAGCACTCCGCCTGGCTGGAGCACCGATGCCGCCCGCCTGCCCTTCGAC  1054
     ||  |||::|||    :::::  |||
379  Leu.SerSerArgAlaValSerAsnPro...............  388

1055 GAGTGGGAACTGACAGCGCGGCCCAAGGCTGATTCGTGGCTGCGGGGCCCAGCA  1104
             ||||||    :::::::     ||||||
389  ...........ProAla  390

1105 CTTGCCTGACGGCCTTGACCTGGCCCTTCAAGGCCCGCAGCCGCGAGGGC  1154
     |||  |||:::  |||    |||||||||  |||
391  .........ProProProAlaIlePro..GlyArgSerAlaProAl  402

1155 TCTTCCTGCTACTGCGCCACTGGGA...CCAAAAACCTGCAGCTACACCTG  1201
     |||||||    |||||||||      :::::|||||||
402  aLeuPro........ProLeuGlyAsnAlaSerArgThrSerThrProP  416

1202 CTGTGCCTACAGCCCAGCGAACCTGTGAAGGCTCCCCCTGCTTGGGC  1251
     |||  ||||||  :::        |||||||||:::|||
416  roVal.ProThrProProSer......LeuProProSerAlaProPr  429

1252 TGGCGCCCCACCAACACTCAAGTCACTGCCGCCCAGGGCTGGCCTCT  1301
     :::  ||||||||    ||||||||      :::|||||||||
429  oSerLeuProProSerAlaProProSerLeuProMetGlyAla.......  443

1302 TGGTGCTGGAAAGTGTAGGCTGGTGCTGCCAGCCTGTCCCCACTGCTTCTT  1351
     ||||||||||||||||
443  .................ProAlaAlaProProLeu........  449

1352 ACTCCCTCCCTGGAGCCCTTTGCCCCCCACAAAAAAGTGCCTGCTGCT  1401
     |||  |||  ::||||||||||||||||      |||  ::  :
321  .....ProLysAlaGlyProAspValLysProAlaValAlaAlaValCysAlaGl  336
```

```
450  ...Pro.ProSerAlaProIleAlaProProLeuProAlaGlyMetProA  465
             ||||||                ||||||
1402 CTCTCCCTCCTCCCACCCACTCACACTCCCCATCCTCTGAGCTC  1451
     ::  |||  ||||||||   :::  |||||||:||||||
465  laAlaProProLeuProProAlaProAlaAlaProProAlaProAla  481

1452 CCTGCAACACAGTGGAAGGTAGAGAGCCACAGTCCCA  1490
     |||||                   |||  :::|||
482  ProAla.................ProAlaAlaPro  487
```

seq_name: pir2:T18535

seq_documentation_block:
high molecular mass nuclear antigen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18535
R;Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch
A;Reference number: Z18955; MUID:9803440
A;Accession: T18535
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1151 <SHI>
A;Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

alignment_scores:
        Quality: 210.50   Length: 509
        Ratio: 0.873   Gaps: 24
Percent Similarity: 47.348   Percent Identity: 25.147

alignment_block:
US-09-786-136-4 x T18535

Align seg 1/1 to: T18535 from: 1 to: 1151

```
202  GTGGTGGGGCTCCACAGACTGCTCCTTGGTTCCACAGGCCATGCCGGA  251
     |||:::||||| |||::|||||
188  ValThrAlaAlaSerGlnSerAlaProLysAlaAlaThrAspAlaAlaAl  204

252  GGTGCTGGCTTCGAGC........................C  268
     |||  |||::|||
204  aValThrAlaAlaSerGlnSerAlaProLysAlaThrValGluValLysP  221

269  CTTCAGACAAGTAAGGTCAGCTTCGCCCCCTCCTGCTTCACTTGCCCTGT  318
     |||::  ||||:::||  :::::: |||::  :::||:::
221  roAla.AlaAlaAlaValAlaLysGluAlaLysGluAlaValThrAlaAlaAl  237

319  AGCTGTGGATGCTACTACTAGGGAGGTCCTGCCCTGGGCCCTGCAG.....  364
     :||::|      ::|||  |||||||  |||:||  :::||
237  aAlaAlaProLysAlaThrAlaGluAlaProValThrSerP  254

365  .......GTCTGCAGCTCAGAGCCCCGCTGAGCCC......CCAGGAACC  400
             |||||||||||||||     |||::||||
254  roThrIleProCysSerSerAlaGluAlaLysProLeuThrAlaAlaSer  270

401  CCTCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCC............  442
     |||  ||||||      :::|||    |||||||||
271  ProThrAlaSerLysAlaThrAlaGluAlaLysProValProAlaThrAl  287

443  .ACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGCGCCT  491
     ::::::|||||||      ::::::::||||||||   :::
287  aSerLeuMetAlaThrLysValThrAlaGluAlaLysProAlaProSerP  304

492  GGGACTTCGGACCCAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCTGCC  541
     :::  :::||||||||||:::  :::::::||||||
304  roSerValProLysAlaThrThrAspThrLysAlaValThrAlaThrAla  320

542  ACTCCCACGGAAGGTCCT.......................  559
     |||  |||
321  .....ProLysAlaGlyProAspValLysProAlaValAlaAlaValCysAlaGl  336
```

1338 .........CCCCCACTGCTTCTTACTCCCTCCCTGGAGCCCTCTTGCCCCC 1379
            ||| :::::|||| ||||| |||| |||||||||||||
600 ysAlaSerProLysAlaLeuMetThrProProProProGlyLeu 616

1380 ACAAAAGTGCCTGCCGTGCTCTCCTCCTCCTCCTCCCACCCACTCACA 1429
     :::::: | ||||||::::::||| ||||||
617 ProArgAlaLeuAlaAlaAlaLysLeuLeuGlyLeuPro........Se 630

1430 CTCCCCTCCATCTTCTGAGCTCCCTGCAACACAGTGGAAGGGTAGAGAGC 1479
     |||||| ||||| ::: || ::: |||
630 rSerProValAlaSerAlaMetHisAlaLysValThrProArgProLeuP 647

1480 -CACAGTCCCCAAATCCTATG 1499
      || ||||:::
647 roAlaSerProValProMet 653

seq_name: pir2:I78557

seq_documentation_block:
N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C;Accession: I78557; I58158; D45219
R;Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.
Neuron 12, 529-540, 1994
A;Title: Developmental and regional expression in the rat brain and functional proper
A;Reference number: I58158; MUID:94206533
A;Accession: I78557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1323 <RES>
A;Cross-references: GB:L31612; NID:g469068; PIDN:AAC37647.1; PID:g469069
A;Accession: I58158
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-66, 'V',68-1323 <RE2>
A;Cross-references: GB:L31611; NID:g469066; PIDN:AAC37646.1; PID:g469067
R;Ishii, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akaz
J. Biol. Chem. 268, 2836-2843, 1993
A;Title: Molecular characterization of the family of the N-methyl-D-aspartate recepto
A;Reference number: A45219; MUID:93155102
A;Accession: D45219
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1265-1323 <ISH>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP;124265)
C;Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology
F;451-879/Domain: glutamate receptor homology <GRH>

alignment_scores:
           Quality: 210.50   Length: 441
           Ratio: 1.341   Gaps: 22
Percent Similarity: 35.601   Percent Identity: 24.717

alignment_block:
US-09-786-136-4 x I78557

Align seg 1/1 to: I78557 from: 1 to: 1323

353 GGGGCCCTGGCAGGTGCAGCTCAGGAGCCGCTGAGCCCCAGGAACCCC 402
     ||| ||| |||||:|||::|||| ||||| ||||| |||||||
889 GlyMetTyrSerCysSerAlaGluAlaProProProAlaLysPr 905

403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCGCCACTGTTGTCT 452
     |||| |||||||:: :||
905 oProPro..............ProProGlnProLeuProSerProAlaT 917

453 AC.. 454
     ||||
917 yrProAlaAlaArgProProProGlyProAlaProPheValProArgGlu 933

560 ....CGCCGCCCACCTCCCCGTCCTCCCGTCGGCTCAGCATGGGCCGCC 605
     ::: |||||| |||||| ||||||
336 uAlaLysProAlaProProGlnProGlnLeuProLys.AlaAla 352

606 GTCACAAACTCTGTAGCCCTGACCCGGGCCAGGCCAACACAGTGAAGGC 655
     :::::: ||| ||||| ::::::
353 AlaAla.............AlaAlaProThrGlyThrGluLe 362

656 AGGGACCATGACTACCTGCCCCTTGGTGGGCTGCAGGAGGCACCAGGCTC 705
     |||||| ||| ||| |::::: ||
362 uLysProAlaThrAlaProProHisGlySerProArgAlaAsnSerHisT 379

706 CTTCCGCTGGACGCGGCCCCTTCGCCCGCCGTGCCATCTCCGAGG... 752
     ::: |||:::||||::::||:::
379 hrValThrValThrProProAsnValProArgAlaAlaAlaThrVal 395

753 ...AGGGCCCTGCGCCGTGCCTCGCCCTTTGCGTGCACCGCGGCCAGCCTC 799
     :::|||||:::::|||||||:: ||||||||:::
396 ProThrAlaGlyAlaValProLysValAlaSerThrGlyThrThrProAlaAl 412

800 AGCC..........................CCACCTCCGC 813
     :||
412 aAlaProGlnGlnLnProValProLysAlaAlaProValThrProProSerP 429

814 CTCATTGCCCTGGGCACTTCTGGCGCCCTGTGTTGGCCAGGGTGACAGTG 863
     ||:::   ::::::: ||
429 roGlnGlnAlaValProArgAlaAlaThrAlaAlaAlaAlaProValThr 445

864 CCACGGCCTCCTGCA...........................GCCCGTCCCCC 889
     |||:: |||::: ||| |||||
446 ProGlnGlnProValThrLysAlaAlaThrThrThrAsnAlaThrProDr 462

890 AGCTGGGGCTCTGAGGGGCCAGGCCAGGTGGACAGTGGGCGGGGCTCAGA 939
     | : ::: ||| ::: :::||||
462 oProGlnProIleProLysAlaAlaAlaProValThrProTHrThrAlaThrProValT 479

940 CACCCAGGGCCTCCGATGGGGCGGAAGGGCTGG...GCGGCCACCGACCTGC 986
     ||::: ||::::   ::: |||||||||
479 hrProGlnGlnProIleProLysAlaGlyThrAspAlaAlaProAlaProPro 495

987 GGGGGCCGGACCTGGGCCACTGGGCGTAGCACTCGCCTGCCTGGAGCACCGA 1036
     ::::: ||:|| ||| ||| :::::
496 AlaValProLysAlaPro.........SerAspGlyArgAlaAlaTh 508

1037 TGCGCCGCTTCGACGAGTGGGAAACTGACAGCGGCCAAGGCTGATTG 1086
     ::||||||||||| ::: :: ||||||||||:::
508 rProGlyValProAsnAlaAlaThrAspProGlnLysProProThr. 524

1087 CTGGCTGCGGCCCCAGCACTTGCCTGACGGCCTTGACCTGGCCGCCCTCA 1136
     :::||||||||:::||| ||
525 ....ProGlnSerValProSerAlaValThrGluPro....... 535

1137 AGGCCCAGCCCGGAGGGCTCTTCCTGCTACTGCCGCCACTCGGACCAAAAC 1186
     ::||||||||| ||||||| |||
536 LysProGln.ProArgAlaAlaProPro.............ProSerA 547

1187 CTGCAGCTACACCTGCTGTGCTACAGCCCAGGAACGTGTGAAGGCTGCC 1236
     ||||||||||||| ::::::|| :::::||
547 snGluAlaThrProAlaValProSerProProAsnLeuLysSer..P 563

1237 CCCTG............................CTGCTTGGGCTGGCCCCCAC 1262
     ||||| ||||||||:::||||:::
563 roLeuProThrIleProValProLeuMetAlaLeuThrProGln 579

1263 CCA...............ACACACTCAAGTCACTGCCCCCCAGGGCT 1294
     ||| ||::: ::: |||||
580 ProValThrAlaGlnMetValThrGlnLeuAlaAlaThrLysPro..... 594

1295 GGCCCTCTTGGTGCTGGGAAAGTGTAGGCTGTGTGCCCAGCCTGTC........ 1337
     :::|||||| ||:::
595 ....SerProIleValProLeu 600

```
1169 .....................TCTAAGGACTTCAGAGAGGCTC
                         ::: |
                         ArgHisC 1171
1225 GTGAAGGCTGCCCCTGCTGCTTGGGCTGCGGCCCCAACACACTCA 1274
     || :::      ||||       ||| |||
     ysAlaSerLeuGluLeuLeu.........ProProProArgHisLeu 1183
1275 AGTCACTGCCGCCCAGGCTGCCGTTGGTGCTGGGAAAGTGTAGGCTG 1324
     |||     ::: |||||          :::
     SerCysSerHisAspGlyLeuAspGly.............GlyTyr 1194
1325 GTGCCAGCCTGTCCCCCACTGCTGTTCTTACTCCCTCCCTGGAGCCCTCTTG 1374
     |  |||    ||| :::||||||||
     pTrpAlaProProPro...........ProProTrpAlaAlaGlyP 1206
1375 CCCCCACAAAAAGTGCCTGCCTGTGCTCTCTCCCTCCCTCCCACCCCAC 1424
     ||    :: |||::| |||   |||||||
     roProProArgArgArgAlaArgCysGlyCysPro......ArgProHis 1220
1425 TCACACTCCCCTCCCATCCTCT 1445
     ||| ::|||  :||:|||
     ProHisArgProAlaSer 1227
```

seq_name: pir1:C45219

seq_documentation_block:
N-methyl-D-aspartate receptor chain NMDAR2D-1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: C45219
R;Ishii, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akaz
J. Biol. Chem. 268, 2836-2843, 1993
A;Title: Molecular characterization of the family of the N-methyl-D-aspartate recepto
A;Reference number: A45219; MUID:93155102
A;Accession: C45219
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1356 <ISH>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:124264)
C;Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology
F;451-879/Domain: glutamate receptor homology <GRH>

alignment_scores:
     Quality: 210.50   Length: 441
     Ratio: 1.341   Gaps: 22
Percent Similarity: 35.601   Percent Identity: 24.717

alignment_block:
US-09-786-136-4 x C45219

Align seg 1/1 to: C45219   from: 1 to: 1356

```
353 GGGGGCCCTGCAGGTGTGCAGCTCAGAGCCCGCCCTGAGCCCCCAGGAACCCC 402
    ||||  |||  |||::::||| ||| |||||||::
889 GlyMetTyrSerCysCysSerAlaGluAlaAlaProProAlaLysPr 905
403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGTTGTCT 452
    |||||||||::: |||||  :::|
905 oProPro..........ProProGlnProLeuProSerProAlaT 917
453 AC................................................ 454
    ||
917 yrProAlaAlaArgProProProGlyProAlaProPheValProArgGlu 933
455 ...........................TCTAAAGGACTTCAGAGAGGCTC 477
                               ::|||||
934 ArgAlaAlaAlaAspArgTrpArgArgAlaLysGly.....ThrGlyPr 948
478 TCCAGCAGGCGCCTGGGACTCGGAC. 502
    ||||::|||||  :::|
948 oProGlyGlyAlaAlaIleAlaAspGlyPheHisArgTyrTyrGlyProI 965
```

```
455 .....................TCTAAGGACTTCAGAGAGGCTC 477
                         ::|||||
934 ArgAlaAlaAspArgTrpArgArgAlaLysGly.....ThrGlyPr 948
478 TCCAGCAGGCGCCTGGGACTCGGAC. 502
    |||::|||||  :::|
948 oProGlyGlyAlaAlaIleAlaAspGlyPheHisArgTyrTyrGlyProI 965
503 ..................CAAAATGCAACTCCAAGCGT 523
                      :::  :::
965 leGluProGlnGlyLeuGlyLeuGlyGluAlaArgAlaAlaProAlaArgGly 981
524 GCTTTGGGGGACCCTGCCACTCCC......ACGGAAGGTCCTCGCCGCCCC 567
    ||| |||  |||||||     |||::||||||:||:|
982 AlaAlaGlyArgProLeuSerProProThrProGlnProGlnLysProP 998
568 ACCTCCCCGTCCTCCCCTGTGGCTCAGCATGGCCGCCGTCACAAACTCT 617
    ||||||
998 oProPro.......................................... 1000
618 GTAGCCCTGACCCGGGCCAGGCCAACACAGTGAAGGCAGCGACCATGAC 667
                                                 :::
1001 ................................Ser 1001
668 TACCTGCCCTTGTGCGCCTGCGAGGACGCA 697
    |||:: ::|||||| ||||||
1002 TyrPheAlaIleValArgGluGlnGluProThrGluProProAlaGlyAlAl 1018
698 ..CCAGGCTCCTTCCGCCTGCGACGGCCCTTCTGCGCCGGCTGTGCGGCA 743
    |||||                    ::|||||:::||
1018 aPheProGlyPheProSerProProAlaProProAlaAlaAlaAlaAlaAlaA 1035
744 TCTCGCAGGAGCCCCTCGCCGT..............GCCTCGCCCTTT 778
    ||||||||              |||||||
1035 laValGlyProProLeuCysArgLeuAlaPheGluAspGluSerProPro 1051
779 GCCGTGACCGGCCAGCCTCAGCCCCACCTCGGCCTCATTGCCCTGGGC 828
    |||   |||          |||  ::|
1052 AlaProSerArg........TrpProArgSerAspProGluSerGlnPr 1065
829 ACTTCTGGGCCCTGGTGTTGGCCAGGTGACAGTGCCACGGCCTCCTGCA 878
    |||||||||| ||:::|||  ||  |||||| |||
1065 oLeuLeuGlyLyGlyAlaGly...GlyProSerAlaGlyAlaProThrA 1081
879 GCCCGTCCCCAGCTCGGGCTCGGGCTCTCGAGGGGCCAGGCCAGTGACAGTGGG 928
    :::|| |||       :::
1081 laProProProArgAlaAlaProAlaProProCysAlaTyrLeuAspLeu 1097
929 GGGGGCTCAGACACCAGGCGCTTCGGATGGGGCGGAAGGGCTGGGCGGCCAC 978
    ::: |||  |||::|||||||||:||||
1098 GluProSerProSerAspSerGluAsp...SerGluSerLeuGlyLyGlyAl 1113
979 CGACCTGCGGGGGGCGGACCTGGGCCACTGCCTAGCACCTCGGCCTGGCTGG 1028
    ::|||||  ||| |||
1113 aSerLeuGlyLyGlyLeuGlu...........ProTrp.Trp 1122
1029 AGCACCGATGCGCCGCTGCCTTCGACGAGTGGGAACTGACACGGCCAAG 1078
    ::|||  ||| ||| :::::
1123 PheAlaSPPheProTyrProTyrAlaGlu.................Ar 1133
1079 GCTGATTGCTGCGTGCGGGGCCCAGCACTTGCCTGACGGCCTTGACCTGGC 1128
    ||||||     |||::|
1133 gLeu.........GlyPro.........................P 1137
1129 CGCCCTCAAGG.............................. 1139
    |||||::|||
1137 roProGlyLyArgTyrTrpSerValAspLysLeuGlyLyTrpArgAlaGly 1153
1140 .............CCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCCAC 1174
                 ||| ::|||  ||| |||||||
1154 SerTrpAspTyrLeuProProArgGlyGlyProAlaTrpHisCys...... 1168
1175 TGGGACCAAAACCTGCAGCTACCACCTGCTGTGCTACAGCCCCAGCGAACGT 1224
```

```
503  ..............................CAAAATGGCAACTCCAAGCGT  523
                                   :::::  :::
965  leGluProGlnGlyLeuGlyLeuGlyGluAlaArgAlaAlaProArgGly  981

524  GCTTTGGGGACCCTGCCACTCCC.....ACGGAAGGTCCTCGCCGCCC  567
     |||  ||| ||| :::|||   |||::  ||||:::||
982  AlaAlaGlyArgProLeuSerProThrProGlnProProGlnLysPr  998

568  ACCTCCCCGTCCTCCCTGTCCGCTCAGCATGGGCCGCCGCCTCACAAACTCT  617
     ||||||||                      ||| :::  |||:||  ||||||
998  OProPro..............................................  1000

618  GTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGGACCATGAC  667
1001 ..............................Ser  1001
                                   :::

668  TACCTGCCCTTGGTCGCGCTGCAGGAGGCA...............  697
     ||||:  ::|||||| ||||||
1002 TyrPheAlaIleValArgGluGlnGlnGluProThrGluProProAlaGlyAl  1018

698  ....CCAGGCTCCTCGCCTGGACGCGCCGCCTTCTTGCCGCCGCTGCCGCA  743
     |||||                       ::|||||
1018 aPheProGlyPheProSerProProAlaProProAlaAlaAlaAlaAlaAlaAla  1035

744  TCTCGCAGGAGGCGCCTCTGCCGT...........GCCTCGCCCTTT  778
     |||||||||                         ||||||
1035 laValGlyProProLeuCysArgLeuAlaPheGluAlaAspGluSerProPro  1051

779  GCCGTGCACCGCGCCCAGCCTCAGCCCCACCTCGGCCTCATTGCCCTGGGC  828
     |||     |||     |||
1052 AlaProSerArg.......TrpProArgSerArgSerProGluSerGlnPr  1065

829  ACTTCTGGGCCCTGGTGTTGGCCAGGTGACAGTGCCACGGCCTCCTGCA  878
     ||| |||||| |||::||||  |||    |||
1065 oLeuLeuGlyGlyLeuGlyAlaGly...GlyProSerAlaGlyAlaProThrA  1081

879  GCCCGTCCCCCAGCTCGGGCTCTGAGGGGCCAGGGCCAGGTGGACAGTGGG  928
     :::|||   |||                ::: |||
1081 laProProProArgAlaAlaProProProCysAlaTyrLeuAspLeu  1097

929  CGGGGCTCAGACACCGAGGCCTCCGATGGGGCGGAAGGGCTGGGCGGCAC  978
     ::: |||  :::|||||| |||    ::|||||||||||||::
1098 GluProSerProSerAspSerGluAsp...SerGluSerLeuGlyGlyAl  1113

979  CGACCTGCGGGGCCGGACCTGGGCCCACTGCCGCTAGCACTCGCCTGGCTGG  1028
     :::|||                       ||| |||  |||
1113 aSerLeuGlyLeuGlu.............GlyProSerAlaGlyAlaProThrA  Pro Trp Trp  1122

1029 AGCACCGATGCGCCGCCGCCTTCGACGAGTGGGAACTGACAGCGGCCAAG  1078
     :::||| |||  ||| :::|||
1123 PheAlaAspPheProTyrProTyrAlaGlu......................Ar  1133

1079 GCTGATTGCTCGCTCGCGGGCCCACTTGCCTGACACTGCCTTGACCTGGC  1128
     ||| ::::||||||||  ::|||||||||||||||::
1133 gLeu................GlyPro.............................P  1137

1129 CGCCCTTCAAG.........................................  1139
     ||||::|||
1137 roProGlyArgTyrTrpSerValAspLysLysGlyGlyTyrArgAlaGly  1153

1140 ...........CCGCAGCCCGGAGGGCTCTTCCTGCTACTCGCGCCAC  1174
                ||| ||||||:::|||
1154 SerTrpAspPheProProArgGlyGlyProAlaTrpProHisCys.....  1168

1175 TGGGACCAAAACCTGCAGCTACACCTGCGTGCTACAGCCCAGGCGAACGT  1224
     :::||| |||||| ||| |||                ||  |||
1169 ....................................ArgHisC  1171

1225 GTGAAGGCTGCCCCCTGCTTGGGCTGCGCCGCCCACCCAACACACTCA  1274
     |||   ||||| |||||||||                  |||:::|||
1171 ysAlaSerLeuGluLeuLeu.................ProProProArgHisLeu  1183
```

```
1275 AGTCACTGCCGCGCCCAGGGCTGGCCTCTTCGTGCTGGGAAAGTGTAGGCTG  1324
     :::   ||||||           ::
1184 SerCysSerHisAspGlyLeuAspGly........................GlyTr  1194

1325 GTGCCAGCCTGTCCCCACTGCGTTCTTACTCCCTCCCTCCTGGAGCCCTCTTG  1374
     ||| ||| ::::|||  |||                   |||||||||:::
1194 pTrpAlaProProPro.........................ProProTrpAlaAlaGlyP  1206

1375 CCCCCACAAAAGTCCTGCTGTCTCCTCCCTCCTCCCTCCCCACCCAC  1424
     |||  :::  ||||:  |||:||  ||
1206 roProProArgArgArgAlaArgCysGlyCysPro.....ArgProHis  1220

1425 TCAACTCCCCTCCATCCTCT  1445
     |||   ||| :::|||
1221 ProHisArgProArgAlaSer  1227
```

seq_name: pir2:S49915

seq_documentation_block:
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C;Accession: S49915
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen-specific genes with extensin-like domains.
A;Reference number: S49915
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

alignment_scores:
```
          Quality:   209.50        Length:     498
          Ratio:       1.012       Gaps:        30
Percent Similarity: 41.566    Percent Identity: 27.711
```

alignment_block:
US-09-786-136-4 x S49915

Align seg 1/1 to: S49915 from: 1 to: 1188

```
239  GGCCATGCCCGGAGGTGCTGGCTTCGAGCCCTTCAGACAAGTAAGGTCAG  288
     ||| |||  :::  ||||||||| |||||:::|||
379  GlyAlaLeuArgProAlaGlnLysThrAlaLeuGlnCysAlaProValIe  395

289  CCTCGCCCCCTCCTGCTTCACTTGCCCGTGAGCTGTGGATGCTACTACTA  338
     ||||||||  :::|||||                    :::|||
395  uAlaArgPro.........................ValAspPro...SerL  403

339  GGGAAGGTCCTGCCTCGGGGCCCTCAGGTGTGCAGCTCAGAGCCCCTGAG  388
     :::||||| ||||||||||||:::
403  ysHisValCysAlaGlyTyrProThrProGlyGlyProProSerSer  419

389  CCC...CCAGGAACCCCT..................CCTGCCTCTCACAG  417
     |||   ||| ||||                     |||:: ||||:||
420  ProValProGlyLysProAlaAlaSerAlaAlaProMetProSerProHisTh  436

418  CCATCTAGATGCAGCTCCTCCACTGTTGTCACTCTAAAGGACTTC  467
     ||| ::::|||:||  |||                  :::|||
436  rProProAspValSerProGluPro..........................LeuP  446

468  AGAGAGGCTTCCAGCAGGCGCCTGGACTCGACCAAAATGGCAACTCC  517
     :::||||||  ||| ||||||:::|||                ||||||
446  roGluProSerProValProAla.ProAlaProMetProMetProThrPr  462

518  A.................AGCGTGCTTGGGGACCCTGCCACTCCCAC  549
     A                 :::  |||  ||||||:::|||:||
462  oHisSerProProAlaAspAspTyrValProProThrProValPro.  478

550  GGAAGGTCCTCGCCGCCGCCACCTCCCCGTCCTCCCTGTCGGCTCAGCATGG  599
```

721 oLysGluProValSerSerProProGlnThrProLysSerSerProP 738

1391 ..CTGCCTGTGCTCTTCCCTCCTCCCCACTCACACTCCCTCC 1438
              |||||| ||||| |||||||||||||:::: ||||||||
738 roAlaProValSerSerPro...ProProThrProValSer.SerProPr 753

1439 A.......   1439

753 oAlaLeuAlaProValSerSerProProSerValLysSerProProP 770

1440 ........TCCTTGAGCTCCCTGCAACACAGTGGAAGGGTAGAGAG 1478
                      ||||||| ||||| ||||| |||
770 roAlaProLeuSerSerProProProAlaProGlnValLys......... 783

1479 CCACAGTCCCCAAATCCTATGCAATAAAGTGCA 1511
        :::||||||| ||||||||| |||:||||
784 ...SerSerProProValGlnValSerSer 793

seq_name: plr2:S27224

alignment_scores:
              Quality: 205.50   Length: 443
                Ratio: 1.309   Gaps: 26
Percent Similarity: 35.440   Percent Identity: 26.185

alignment_block:
US-09-786-136-4 x S27224

Align seg 1/1 to: S27224   from: 1   to: 1323

353 GGGGCCCTGCAGGTGTGCAGGTCAGGAGCCCGTGAGCCCCAGGAACCCC 402
              ||||||::: ||||||||| |||||::: ||
889 GlyMetTyrSerCysCysSerAlaGluAlaAlaProProAlaLysPr 905

403 TCCTGCCTCTCCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGTTGTCT 452
              |||| |||||||||:::: ||||
905 oProPro.........   ProProGlnProLeuProSerProAlaT 917

453 ACTCTAAAGGACTTCAGAGAGAGCCTCTTCCAGCAGGCGCCCTGGGACTCGGAC 502
              || ||||||||
917 yr................   ProAla.................. 919

503 CAAAATGGCAACTCCAAGCGTGCTTTGGGGACCCTGCCACTCCACCGGA 552
                                            ||| |||
920 ....   AlaArgProProPr 924

553 AGGTCCTCCGCGCCACCTCCCCGTCCTCCCGTCGCGCTCAGCATGGGCC 602
              |||||| ||||| :::
924 oGlyProAlaProPheValProArgGluArgGlnAlaAlaAlaAspPArgTrpA 941

603 GCCGTCACAAACTCTGTAGCCCTGACCCGGCACCCGGCCCAACAACAGTGAA 652
              ||| ||| ||| :::::||
941 rgArgAlaLysGlyThrGlyPro...ProGlyGlyAlaAlaLeuAlaAsp 956

479 GlyLysSerProProAlaThrSerProSer..ProGlnValGlnProPro 494

600 GCCGCCGTCACAAACTCTGTAGCCCTGACCCGGGCCAGGCCAACAACAGT 649

495 AlaAlaSerThrProProSerLeuValLysSerProProGlnAl 511

650 GAAGGCAGGCGACC.......ATGACTACCTGCCCTTGGTGCG 684

511 aProValGlySerProProProValLysThrThrSerPro....... 525

685 GCTGCAGGAGGAGCCACCAGGCTCCTTCCGCCTGGACGCGCCCTTCTGCGCCG 734

526 ....ProAlaProIleGlySerProSerProProProPro 537

735 CTGTGCCGATCTCCGAGGAGCCGCCTCTGCCGTGCCCTCGCCCTTG..... 779

538 Val..............SerValValSerProProProProValLysSe 549

780 ....CCGTCGCCGCCGCCAGCCTCAGCCCCACCTCGGCCTCAT....... 818

549 rProProProAlaProValGlySerProProProProGluLysSerP 566

819 .....TGCCTGGGCACTTCTGGGCCCCTGTGTTGGCCAGGGTGACAGTG 863

566 roProProPro............AlaProValAlaSer 574

864 CCACGGCTCCTGCAGCCCGTCCCCAGCTCGGGCTCTGAGGGGCCAGGC 913

575 ProProProProValLysSerProProProProThrLeu.......... 587

914 CAGGTGGACAGTGGCGGGCTCAGACACCAGGCTCCGATGGGGCGGA 963

588 ....ValAlaSerProProProVal....... 595

964 AGGGCTGGGCGGCCACCGACCTGGGCGGCCACTGCCCGTAG 1013

596 ....LysSerProProProAlaProVal 604

1014 CACTCCGCTGGCTGGAGCACCGATGCGCGCGTGCCTTCGACGAGTGGGAA 1063

605 AlaSerProProProValLysSerProProProThrProValAl 621

1064 CTGCACAGCGGCCAAGGCTGATTGCTGCGTGCGGGCCCAGCACTTGCCTGA 1113

621 aSerProProProProAla.ProValAlaSerSerProProMetLys 637

1114 CGGGCCTTGACCTGGCCGCCCTCAAGGCCCCAGGCCCGGAGGGCTCTTCCTGC 1163

638 SerProProProThrProThrProValSerSerProProPro.....ProGl 652

1164 TACTGCGCCCACTGGACCAAAACCTGCA...GCTACACCTGCTGTGCTAC 1210

652 uLysSerProProProProProProAlaLysSerThrPro.......... 665

1211 AGCCCAGGCAAGCTGTGAAGGCTGCCCCCTGCTTGGGCTGGCGCCCC 1260

666 ..ProProGluGluTyrProThr.ProProThrSerValLysSerSerPr 681

1261 ACCCAACACTCAAGTCACTGCGCCCCAGGGCTGGCCTTCTTGGTGCTGG 1310

681 oProProGlu..LysSerLeuProProProThr............... 691

1311 GAAAGTGTAGGCTGGTGCCCAGCCTGTCCCCACTG............... 1345

692 ....LeuIleProSerProProProGlnGluLysProThrPro 704

1346 CTTCTTACTCCCCCCTGGAGCCCTTGCCCCCACAAAAAGTGC..... 1390

705 ProSerThrProSerLysProProSerProGluLysProSerProPr 721

1390 1390

```
                                         |||||||||||:::  |
                                         ProProTrpAlaAlaGlyP 1206
1194 pTrpAlaProProPro...........
     |||  |||
1369 CTCCTTGCCCCCACAAAAAGTGCCTGCCTGTGCTCTCCCTCCCTCCTCCCA 1418
     ||  |||||||   ::::|||  |||
1206 roProAlaProArgArgAlaArgCysGlyCysPro..........Arg 1218

1419 CCCCACTCACACTCCCCTCCATCCTCT 1445
     ||||| |||  |||  :::||
1219 ProHisProHisArgProArgAlaSer 1227
```

```
653  GGCAGCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGG 702
     |||     :::  |||   |||  ||
957  GlyPheHisArgTyrTyrGlyProIle..........GluProGlnGly 969

703  CTCCTTCCGCCTGGACGCGCCCTTCTGCCGCCGCTGTGCCCATCCGCAGG 752
     |    :::|||  ||:::||||||
969  yLeuGlyLeuGlyGluAla.ArgAlaAlaPro..........Arg 980

753  AGGGCCCTCGCGTGCCTCCGCCCTTGCCCTGCACGCGGCCAGCCTCAGC 802
     ::|||:::|::  |||   |||  |||  :::
981  GlyAlaAlaGlyAlaProLeuSerProProThrGlnProProGlnLy 997

803  CCCACCTCGGCCTCATTGCCCTGGGCACTTCTGGGCCCTGGTGTTGGCCA 852
     |||||||        :::||||:::|||   |||:::|||  :
997  sProPro..........ProSerTyrPhe..AlaIleValArgGluG 1009

853  GG..........GTGACAGTG 863
     ::
1009 lnGluProAlaGluProProAlaGlyAlaPheProGlyPheProSerPro 1025

864  CCACGGCCTCCTGCAGCC.................. 881
     |||  |||||||||||
1026 ProAlaProProAlaAlaAlaAlaAlaAlaAlaValGlyProLeuCysAr 1042

882  .................CGTCCCCAGCT...........CGGGCTC 900
     |||||||||          :::|||
1042 gLeuAlaPheGluAspGluSerProProAlaProSerAlaGlyArgValL 1059

901  TGAGGGGCCAGGCC...AGGTGGACAGTGGGCGGG...........GCT 935
     ||    ::::||  |||   ||||||||
1059 euThrProArgAlaSerArgCysTrpValGlyAlaArgAlaAlaArgAla 1075

936  CAGACACCGAGGCCTCCGATGGGGCGGGAAGGGCTGGGCGGCACCGACCTG 985
     ||||||||||      |||  |||  || ||
1076 LeuGlyProArgProHisHisArgValArgThrAlaProProProCys 1092

986  C.................. 986
     :
1092 sAlaTyrLeuAspLeuGluProSerProSerAspSerGluSerGlus 1109

987  .........GGGGCCGGACCTGGGCCACTGCCGTAGCACTCGCCTGGCTGG 1028
     |||||||:::  |||  |||     ::::|||
1109 erLeuGlyLeuGlyAlaSerLeuGlyLeu.......GluProTrpTrp 1122

1029 AGCACCGATGCGCCCGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAG 1078
     ::||| |||  |||  ||  :::|||
1123 PheAlaAspPheProTyrProTyrAlaGlu................Ar 1133

1079 GCTGATTGCTGCTGCGGGCCCAGCACTTGCCTGACGCGGCCTTGACCTGGC 1128
     |||||||||       |||   |||  ||
1133 gLeu.........GlyPro..........P 1137

1129 CGCCCTCAAGG.................. 1139
     |||||||:::|||
1137 roProGlyArgTyrTrpSerValAspLysLeuGlyGlyTrpArgAlaGly 1153

1140 ..........CCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCAC 1174
     |||   |||   ::|||  :::|||
1154 SerTrpAspTyrLeuProProArgGlyGlyProAlaTrpHisCys..... 1168

1175 TGGGACCAAAACCTGCAGCTACACCTGCTGTGCTACAGCCAGCGAACGT 1224
     ::||| |||  |||   ::::||
1169 ..........ArgHisC 1171

1225 GTGAAGGCTGCCCCGTGCTGCTTGGGCTGGCGGCCCCACCCAACACACTCA 1274
     ||  :::   |||||||     |||  |||  |||
1171 ysAlaSerLeuGluLeuLeu.........ProProProArgHisLeu 1183

1275 AGTCACTGCCGCCCAGGGCTGGCCTTTGGTGTCTGGGAAAGTGTAGGCTG 1324
     |||    :::  ||||||
1184 SerCysSerHisAspGlyLeuAspGly...........GlyTr 1194

1325 GTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCCTGGAGC.......C 1368
```

```
OM of: US-09-786-136-4 to: SwissProt_40:*    out_format : pfs
Date: Sep 19, 2002  5:46 AM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US0978616/runat_16092002_141245_19270/app_query.fasta_1.1592
 -DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -ANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -X_MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
 -RM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09786136 @CGN_1_22 -NCPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-136-4
Query length: 1511
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 37.740000
```

```
SwissProt_40:CA11_CHICK  - 173.50 153.44 0.0516 1453 ! P02457 gallus gallus (ch
SwissProt_40:APXL_HUMAN  + 173.50 152.71 0.0510 1616 ! Q13796 homo sapiens (hum
SwissProt_40:CA21_HUMAN  - 171.00 151.68 0.0687 1366 ! P08123 homo sapiens (hum
SwissProt_40:CA11_CANFA  - 170.50 150.79 0.0721 1460 ! Q9xsj7 canis familiaris
SwissProt_40:CA21_ONCMY  - 170.00 150.86 0.0769 1356 ! O93484 oncorhynchus myki

seq_name: SwissProt_40:T2D3_HUMAN

seq_documentation_block:
ID   T2D3_HUMAN     STANDARD;       PRT;   1083 AA.
AC   O00268; Q99721; Q9BX42; Q9BR40;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
DE   (TAFII135) (TAFII-130) (TAFII130).
GN   TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97336072; PubMed=9192867;
RA   Mengus G., May M., Carre L., Chambon P., Davidson I.;
RT   "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT   of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT   mammalian cells.";
RL   Genes Dev. 11:1381-1395(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaslaiho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [3]
RP   SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=97098442; PubMed=8942982;
RA   Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT   "Molecular cloning and analysis of two subunits of the human TFIID
RT   complex: hTAFII130 and hTAFII100.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC   -!- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC       PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC       ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC       BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC   -!- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC       NUMBER OF TBP-ASSOCIATED FACTORS (TAFS).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
```

```
score_list:
Sequence                 Strd  Orig  ZScore EScore Len  | Documentation
SwissProt_40:T2D3_HUMAN   + 240.50 213.81 3.0e-05 1083 ! O00268 homo sapiens (human)
SwissProt_40:GP1_CHLRE    + 235.00 213.62 6.0e-05  555 ! Q9fpq6 chlamydomonas reinhard
SwissProt_40:YHL1_EBV     - 218.50 198.05 0.0004   660 ! P03181 epstein-barr virus (str
SwissProt_40:EXTN_TOBAC   + 214.00 194.57 0.0006   620 ! P13983 nicotiana tabacum (co
SwissProt_40:NME4_MOUSE   + 211.50 187.18 0.0007  1323 ! Q93391 mus musculus (mouse
SwissProt_40:NME4_RAT     + 210.50 186.30 0.0008  1323 ! Q62645 rattus norvegicus (rat
SwissProt_40:EBN1_EBV     - 205.00 186.50 0.0017   641 ! P03211 epstein-barr virus (str
SwissProt_40:NME4_HUMAN   + 200.00 177.09 0.0027  1336 ! O15399 homo sapiens (human
SwissProt_40:TEGU_EBV     + 199.50 170.76 0.0026  3149 ! P03186 epstein-barr virus (st
SwissProt_40:YHL1_EBV     + 196.50 178.89 0.0043   660 ! P03181 epstein-barr virus (str
SwissProt_40:MUC2_HUMAN   + 196.50 168.73 0.0034  5179 ! Q02817 homo sapiens (human
SwissProt_40:VNUA_PRVKA   + 192.00 168.34 0.0064  1733 ! P33485 pseudorabies virus
SwissProt_40:SYN1_CANFA   + 191.50 177.30 0.0085   415 ! Q62732 canis familiaris (dog
SwissProt_40:CA12_MOUSE   + 189.50 167.34 0.0086  1459 ! P28481 mus musculus (mouse
SwissProt_40:CA17_HUMAN   - 189.50 162.51 0.0080  2944 ! Q02388 homo sapiens (human
SwissProt_40:CA12_HUMAN   + 188.50 166.67 0.0097  1418 ! P02458 homo sapiens (human
SwissProt_40:SYN1_HUMAN   + 186.00 169.30 0.0139   705 ! P17600 homo sapiens (human
SwissProt_40:PROP_HUMAN   + 182.00 168.62 0.0228   469 ! P27918 homo sapiens (human
SwissProt_40:MY15_HUMAN   + 182.00 154.73 0.0180  3530 ! Q9ukn7 homo sapiens (human
SwissProt_40:SYN1_BOVIN   + 181.50 165.37 0.0230   706 ! P17599 bos taurus (bovine).
SwissProt_40:VGLG_HSV2H   + 181.00 165.00 0.0243   699 ! P13290 herpes simplex virus
SwissProt_40:WAIP_HUMAN   + 180.50 166.83 0.0268   503 ! O43516 homo sapiens (human
SwissProt_40:SYN1_RAT     + 180.50 164.52 0.0257   704 ! P09951 rattus norvegicus (rat
SwissProt_40:HFC1_HUMAN   + 180.00 156.78 0.0240  2035 ! P51610 homo sapiens (human
SwissProt_40:SE5_RAT      - 179.00 162.12 0.0298   825 ! Q63003 rattus norvegicus (rat
SwissProt_40:NTC3_MOUSE   + 179.00 155.01 0.0264  2318 ! Q61982 mus musculus (mouse
SwissProt_40:CA21_CHICK   + 178.50 158.24 0.0298  1362 ! P02467 gallus gallus (chick
SwissProt_40:WAS1_HUMAN   + 177.50 163.49 0.0370   559 ! Q92558 homo sapiens (human
SwissProt_40:SYN1_MOUSE   + 177.50 162.34 0.0362   670 ! O88935 mus musculus (mouse
SwissProt_40:PRP3_MOUSE   + 177.00 167.43 0.0421   296 ! P05143 mus musculus (mouse)-
SwissProt_40:SPD1_NEPCL   + 177.00 161.06 0.0378   747 ! P19837 nephila clavipes (orb
SwissProt_40:T2D3_HUMAN   - 177.00 158.51 0.0361  1083 ! O00268 homo sapiens (human
SwissProt_40:DRPL_RAT     + 176.50 157.46 0.0378  1183 ! P54258 rattus norvegicus (rat
SwissProt_40:CA11_MOUSE   - 175.50 155.18 0.0413  1453 ! P11087 mus musculus (mouse)
SwissProt_40:CA13_MOUSE   - 174.50 154.26 0.0461  1464 ! P08121 mus musculus (mouse
SwissProt_40:TEGU_HSV11   + 174.50 148.95 0.0421  3164 ! P10220 herpes simplex virus
SwissProt_40:CA11_RAT     - 174.00 159.19 0.0535   671 ! P02454 rattus norvegicus (rat
SwissProt_40:CA12_BOVIN   - 174.00 158.45 0.0528   747 ! P02459 bos taurus (bovine).
SwissProt_40:GSR1_HUMAN   - 174.00 153.61 0.0486  1509 ! Q9nzm4 homo sapiens (human
SwissProt_40:CA21_BOVIN   - 173.50 153.87 0.0520  1364 ! P02465 bos taurus (bovine)
```

(handwritten annotation: "Not Prev relevant")

```
DR   EMBL; Y11354; CAA72189.1; -.
DR   EMBL; AL137077; CAC36006.1; -.
DR   EMBL; AL109911; CAC22312.2; -.
DR   EMBL; U75308; AAC50901.1; -.
DR   HSSP; P19656; 1AFH.
DR   TRANSFAC; T02328; -.
DR   MIM; 601796; -.
DR   InterPro; IPR003894; TAF_hom.
DR   SMART; SM00549; TAFH; 1.
KW   Transcription regulation; Nuclear protein.
FT   DOMAIN        39     42       POLY-HIS.
FT   DOMAIN        52     57       POLY-ALA.
FT   DOMAIN        98    101       POLY-GLY.
FT   DOMAIN       142    148       POLY-ALA.
FT   DOMAIN       268    275       POLY-PRO.
FT   DOMAIN       331    337       POLY-ALA.
FT   DOMAIN       680    683       POLY-PRO.
FT   DOMAIN       808    813       POLY-ALA.
FT   DOMAIN       828    831       POLY-ASP.
FT   CONFLICT     105    117       PGPPSPRRPLVPA -> GRGLLQQRGGRES
FT                                 (IN REF. 3).
FT   CONFLICT     136    136       A -> S (IN REF. 2).
FT   CONFLICT     185    185       G -> GPG (IN REF. 2).
FT   CONFLICT     233    264       MISSING (IN REF. 3).
FT   CONFLICT     293    293       P -> L (IN REF. 3).
SQ   SEQUENCE   1083 AA;  109943 MW;  A6453B27572A0752 CRC64;
```

```
alignment_scores:
     Quality:   240.50                Length:    452
     Ratio:       1.221               Gaps:       26
Percent Similarity: 43.584       Percent Identity: 27.434

alignment_block:
US-09-786-136-4 x T2D3_HUMAN    ..

Align seg 1/1  to: T2D3_HUMAN  from: 1 to: 1083
```

```
338 AGGGAGGTCCTGCCTGGGCCCTG......CAGGTGTGCAGCTCAGAGCC 381
    |||  :::||||||||      :::||| :::|||  ||||:::::||
 51 ArgAlaAlaAlaGlyAlaLeuGlyAsnHisValSerGlySerPr 67

382 CGCTGAGCCCCCAGGAACCCCCTCCTGCCTCTCACAGCCATCTAGATGCAG 431
    |||   |||::: ||| ||||||   :::|||||| ::::|||
 67 oAlaGlyAlaAlaGlyAlaGlyPro..............AlaA 77

432 CTCCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCA 481
    |||||   ::||::|||||| :::|||:::||||
 77 laPro..............AlaGluGlyAlaPro 83

482 GCAGGCGCCTGGGACTCGGACCCAAAATGGCAACTCCAAGCGTGCTTTGGG 531
    ::::|||   ||| |||||||| :::  :||
 84 GlyAlaAlaProGluProProAlaGlyArgAlaArgProGlyGlyGl 100

532 GGACCCTGCCACTCCC...ACGGAAGGTCCTCGCCGCCCA........ 568
    | ||  :::::||| ::::|||||||||
100 yGlyProGlnArgProGlyProProSerProArgProLeuValProA 117

569 .......CCTCCC........CGTCCTCCCTGTCGGCTCAGC 595
    |||||    ||||||||  :::  |||
117 laGlyProAlaProProAlaAlaLysLeuArgProProGluGlySer 133

596 ATGGGCCGCCGTCACAAACTCTGTAGCCCTGACCCGGCCAGGCCAACAA 645
    |||:::|||   |||:::|| |||:::  |||
378 GlnGlyAlaLeuProSerProAlaAlaValPro......ProProAlaPr 392
```

```
134 AlaGly..............AlaCysAlaProValProAlaAlaAlaAlaAlaVa 146
646 CAGTGAAGCAGCGACCATGACTACCTGCCCTGGTCGCGGCTGCAGGAGG 695
    :::  |||    :::
146 lAlaAlaGlyProGlu.................................. 151
696 CACCAGGCTCCTTCCCCTGACGCGGCCCCTTCCTGCGCCGCCGTGTGCGCATC 745
    ||| ||||||  ||:::||| ||::::||||:::|||
152 ..Pro.AlaProAlaGlyProAlaLysProAlaGlyProAlaAlaLeuAl 167
746 TCGCCAGGAGGCCGCCTGCCGTCGCCCCTTTGCCGTCGCCACCGCCGCCAG 795
    :::|||    ||| ||| ||||:::|||
167 aAlaArg.......AlaGlyProGlyProGlyProGlyProGlyProG 181
796 CCTCAGCCCCACCTCGGCCTCATTGCCCTGGG.................... 827
    ::: ::|||   :::|||   ||||||
181 lyProGlyProGlyLysProGlyAlaGlyAlaGlnThrLeuAsn 197
828 ..................CACTTCTGGGCCCTGGTTGGCCAG 853
    |||  ||
198 GlySerAlaAlaLeuLeuAsnSerHisHisAlaAlaAlaProAlaValSe 214
854 GGTGAGTGCCACGGCCTCTGCAGCCGTCCC...CCAGCTCGGGCTC 900
    ::::::   |||  |||||| :::
214 rLeuValAsnAsnGlyProAlaAlaLeuProLeuProLysProAlaA 231
901 TGAGGGGCCAGGCCAGTGGACA.........GTGGGCGGGGCTCAGACA 941
    ||||||||:::||| |||||||:::|||
231 laProGlyThrValIleGlnThrProProPheValGlyAlaAlaAlaPro 247
942 CCGAGGCCTCCGATGGGGCGGAAGGGCTGGGCGGCACCGACCTGCGGGGC 991
    ||||||||||| :::
248 ProAlaPro............AlaAlaProSerProProAl 257
992 CGGACCTGGGCCACTGCCGTAGCACTCGCCTGGCTGGAGCACCGATGCGC 1041
    :::||||||:::|||
257 aAlaProAlaProAla........AlaProAlaAlaProProProP 271
1042 CGCTGCCTTGCGACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGGC 1091
    ||  ||:::
271 roProProAla............................... 274
1092 TGCGGGCCCAGCACTTGCCTGACGGCCTTGACCTGCCGCCCTCAAGGCC 1141
    ||||:::|||   |||||| ||||:::|| |||||
275 ......ProAlaThrLeuAla.ArgPro...ProGlyHisProAlaGly. 287
1142 GCAGCCCGAGGGCTCTTCCTGCTACTGCGCCA.................. 1173
    |||  ||| |||||||:::::
288 ..ProProThrAlaAlaProAlaValProProProAlaAlaGlnAsn 303
1174 ........CTGGGACCAAAACCTGCAGCTACACCTGCTGCTACAGCC 1214
    ||| |||||| :::|||||  :::  :::||
304 GlyGlySerAlaGlyAlaAlaProAlaProAla.AlaGlyGlyP 320
1215 CAGCGAACGTGTGAAGGCTGCC...CCTGCTGCTGGGCTGGCCGCCCCA 1261
    ||||||:::|| |||||||| :::||||||
320 roAlaGly:::ValSerGlyGlnProGlyProGlyAlaAlaAlaAlaPro 336
1262 CCCAACACACTCAAGTCACTGCGCCCAGGGCTGGCCCTCCTTGGTGCTGGG 1311
    ::: |||  :::
337 AlaProGlyValLysAlaGluSerProLys.................. 346
1312 AAAGTGTAGGCTGGTGCCAGCCTGTCCCCA.................... 1342
    ||:::||| ||||||||||||
347 ......ArgValGlnAlaAlaProProAlaAlaGlnThrLeuAlaA 361
1343 ....................CTGCTTCTTACTCCCCTCCCTG 1363
    |||:::||||||||||
361 laSerGlyProAlaSerThrAlaAlaSerMetValIleGlyProThrMet 377
1364 GAGCCCTTCTTGCCCCCACAAAAAGTGCTGCCGTGTCCTCTCCCCTCC 1413
    :::  ||| |||:::||||
378 GlnGlyAlaLeuProSerProAlaAlaValPro......ProProAlaPr 392
```

us-09-786-136-4.rsp

Percent Similarity: 39.950   Percent Identity: 26.633

alignment_block:
US-09-786-136-4 x GP1_CHLRE

Align seg 1/1 to: GP1_CHLRE from: 1 to: 555

```
 323 GTGGATGTACTACTAGGGAGGTCCTGCCTGGGCCCTGCCAGGTGTCCAG  372
      |||  :::::::::: ::: :::||||||:::::||  |||
  24 ValAlaSerAlaasnAlaGlnCysValProGlyGlyIlePheAsnCysPr   40

 373 CTCAGAGCCCGCTGAGCCCCCAGGAACCCTCTGCCTCCTCCACAGCCATC  422
      ::|||||| |||    ::|||||
  40 oProSerProAlaProSerProAlaProPro.................   51

 423 TAGATGCAGCTCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGA  472
      :::||| ||
  52 .........SerProAlaProPro.......................   56

 473 GGCTCTCCAGCAGGCGCCTGGGACTCGGACCAAAAATGGCAACTCCAAGCG  522
      |||||||||
  57 ...SerProAlaProPro.......................   61

 523 TGCTTTGGGGACCCTGCCACTCCCACGGAAGGTCCTCGCGCCCA...C  569
      ::|||||| |||::: ||||||  ||| |
  62 ..........SerProAlaProSerProGlyProProSerProAlaP   75

 570 CTCCCCGTCCTCCCTGTCGGCTCAGCATGGGCGCCGTCACAAACTCTGT  619
      ||||| ||||||
  75 roProSerProProSerProAlaProPro.................   84

 620 AGCCTGACCGGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGACTA  669
      |||||| ||| |||::  ||  ||
  85 SerProAlaPro.ProSerProAlaProProSerProAlaProProSerP  101

 670 CCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACG  719
      ::|||  |||::   |||||||| ::: :::
 101 roAlaProProSerProAlaProProSerProSerProAlaProAlaP   117

 720 CGCCCTTCTTGCGCGCCGCTGCGCATCTCGCAGGAGCGGCCTCTCGCCGTGCC  769
                                            ::: SerProAlaProPr  127
 118 ProProSerProPro.................

 770 TCGCCCTTGCCGTGCACCGGCAGCCTCAGCCCCACCTCGGCCTCATT  819
      | ||| |||::: |||::: ::|||||| ||
 127 oSerProSerProAlaProProSerProSerProAlaProAlaP   144

 820 GCCCTGGGCACTTCTGGGCCTGGTGTTGCCAGGTGACAGTGCCACGG  869
      |||                  ||
 144 roPro...........LeuProProSerProAla   151

 870 CCTCCTGCAGCCCGTCCCCCAGCTCGGGCTCTGAGGGGGCCAGGCCAGGGTG  919
      ||||||::: |||||||
 152 ProProSerProProVal.................   159

 920 GACAGTGGGCGGGGCTCAGCACCGAGGCCTCCGATGGGCGGAAGGGCT  969
      ::|||  |||||||  ::
 160 ..........ProProSerProProValProSer.....   170

 970 GGGGCGCACCGACCTCGGGGGCCAGACCTGGCCACTGCGTAGCACTCG  1019
      |||||  |||::::||| |||
 171 ..ProAlaProProSerProThrProProSerProSerProValPro  186

1020 CCTGGCTGAGCACCAGCATCGCCCGCTGCCTTCGACGAGTGGAACTGACA  1069
      |||::: |||||| |||::::|||
 187 ProSerProSerProAlaProAlaProProSerProAlaProProSerProAla   203

1070 GCGGCCAAGGCTGATTGCTGGCTGCGGGGGCCCAGCACTTGCCTGACGGGCCT  1119
      |||||  |||::
 203 aProProSerProAlaProProValProProSerProAla.........   216

1120 TGACCTGGCCGCCCTCAAGGCCGCAGCCCCAGG.....GCTCTTCCTGC  1163
      ||||||||| ||| ||||||           ::: |||::::
```

```
                          Ratio:    0.967    Percent Identity:  38.567    Gaps:        34
Percent Similarity:   38.567    Percent Identity:  26.792

alignment_block:
US-09-786-136-4/rev x YHl1_EBV    ..

Align seg 1/1 to: YHl1_EBV   from: 1   to: 660

 217 .......ProProSerProPro.SerProAlaProProSerProProSe 230       1481 TGGCTCTCTACCTTCCCACTGTGTCGCAGGAGCTCAGAGGATGGAGAGGG 1432
1164 TACTGCGCCACTGGACCAAAACCTCAGCTACACCTGCTGTGCTACAGC 1213          ||| :::       ||| ||:::|| |||
  : |||||| ::||||:::|||| ||||::||| ::|::  |||::|         148 TrpSerAlaArgAsnProGlyCysProArgThr................ 158
 230 rProAlaProProSerProSerProProAlaProProSerProValProP 247

1214 CAGCGAACGTGTGAAGGCTGCCCCCTGCTGCTTGGGCTGGCGCCCCACC 1263      1431 AGTGTGAGTGGGTGGGAGGAGAGGAGAGGACACAG............... 1395
 |||||           ||| ||||||:::            ||||||||              ||||:::|:::: ||||| ||||||
 247 roSer......ProAla.ProProSerPro.......AlaProPr 257            159 ...........TrpArgArgSerGlyAlaGlnArgGlyHisProP 171

1264 CAACACACTCAAGTCACTCGCCGCCAGGGCTGGCCCTCTTGGTGCTGGGAA 1313     1394 .................GCAGGCACTTTTGTGGGGCAAGAGG 1368
 |:::|||                                   ||||||                        ::|||
 257 oSerProLysProSerProAlaProPro................... 265           171 roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 187

1314 AGTGTAGGCTGGTGCCAGCTGTCCCCACTGCTTCTTACTCCCCTCCCTG 1363      1367 GCTCCAGGGAGGGAGTAAGAGCAGTGGGGACAGGCTGGCACCA....... 1323
 |||||           ||||||                  |||||||                ||||||||              :::||||
 266 ...........ProPro..........ProSerPro 270                     188 AlaProGly...........AlaProGlyThrProAlaAlaProGlyPr 200

1364 GAGCCCTTGCCCCACAAAAAGTGCCTGCCTGTCTCTCCCTCC 1413            1323 .................................................. 1323
 |||            ::        :::|||:::||| ||
 271 ProProProProProArgProProPheProAlaAsnThrProMetPr 287         200 oGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgG 217

1414 TCCCACCCCACTCACACTCCCCTCCATCCTCGAGCTCCCTGCAACACAG 1463      1322 ...........GCCTACACTTTCCCAGCACCAAGA 1299
 |||::||||     ||||||||||   ||| |||  |||                                   ::||| |||
 287 oProSerPro.......SerProProSerProAlaProProThrPro 302          217 lySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArg 233

1464 TGGAAGGGTAGAGAGCCACAGTCCCAAATCCTATG 1499                     1298 GGCCAGCCCTG....................GCGGCAGTGACTTGAGTGTGTTGGT 1261
 ::|||:::|| |||:::||||:::||                                    ::|||||              ||| ||    ::
 303 ProThrProProSerProSerProVal 314                             234 GlnGluProArgLeuProGlnAspProAlaLeuAlaAlaGlnArgCysProAl 250

seq_name: SwissProt_40:YHl1_EBV                                 1260 GGGGGCCCAGCCCAAGCAGCAGCAGGGGCAGCC.................. 1230
                                                                 :||| |||         :::|||||||
seq_documentation_block:                                         250 aGlyProProThrArgSerGlyAlaAlaAlaGlaAlaArgThrHisArgA 267
ID   YHl1_EBV      STANDARD;      PRT;       660 AA.
AC   P03181;                                                    1229 .................TTCAACACGTTCGCTGGGCTGT.......... 1209
DT   21-JUL-1986 (Rel.  01, Created)                                               |||||:::|||||||
DT   21-JUL-1986 (Rel.  01, Last sequence update)                267 rgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 283
DT   15-JUL-1998 (Rel.  36, Last annotation update)
DE   Hypothetical BHLF1 protein.                                 1208 .......AGCACAGCAGTGTAGCTGCAGGTTTTGGTCCCAGTGGCGCAGT 1165
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).                 ::||||||:          ||| ||    |||||||
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;        284 TrpArgArgSerGlyAlaGlnArgGlyHisProProProGlyAla. 299
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;                                           1164 AGCAGGAAGAGCCCTCGGGCTGCGGCCTTGAGGGCGGCCAGTCAAGGGCC 1115
RN   [1]                                                          |||||||||         |||||||| ||| |||  |||||
RP   SEQUENCE FROM N.A.                                          300 ...........GlyGlnArgProSerGly...ProThrGlyGlyA 310
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,   1114 GTCAGGCAAGTGCTGGG..............CCCGCAGCCAGCAATCAGCCT 1077
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequin C.,     || ||||:::   ||||||              ||||||||||||  ::
RA   Tuffnell P.S., Barrell B.G.;                                310 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).                                   1076 TGGCCGTGTGCAGTTCCCACTCGTCGAAGGCAGCAGCGGCCATCGGTGCT.. 1029
                                                                 ::|||||||||         |||||||||| ||| |||
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration   327 GlyGlyAlaAlaValProSer..............GlyAlaTh 336
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its   1028 ......CCAGCCAGGCAGTGCTACGGCAGTGGCCCAGGTCCGGCGCCCCG 986
CC   use by non-profit institutions as long as its content is in no way               ||||||||||||            |||||||||||||
CC   modified and this statement is not removed. Usage by and for commercial   336 rProHisProGluArg.................GlySerGlyProA 346
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).                    985 CAGGTCGTGCCGCCCAGCC...CTTCCGCCCATCGGAGG.......... 948
                                                                 ||              ||||||:::::|:::|
DR   EMBL; V01555; -; NOT_ANNOTATED_CDS.                         346 laAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArg 362
DR   PIR; A03742; Q0BE3.
KW   Hypothetical protein; Early protein; Repeat.                947 ...CCTCGGTGCTGCGAGCCCCGCC...CACTGTCCACCTGGCCTGGCCCC 904
FT   DOMAIN        149     648    4 X 125 AA TANDEM REPEATS.      ||::::    |||:::  ::||||||  |||           ||      Pr 377
FT   REPEAT        149     273    1.                             363 LeuProGlnAspProLeuAlaAlaAlaGlnArgCysProAlaGly......
FT   REPEAT        274     398    2.
FT   REPEAT        399     523    3.                             903 TCAGAGCCCCAGCTGGGGACGGGCTGC...............AGGAGGCCGT 866
FT   REPEAT        524     648    4.
SQ   SEQUENCE     660 AA;  66244 MW;  86DA1D67A37152A2 CRC64;

alignment_scores:
     Quality:   218.50              Length:       586
```

632 erGlyAlaAlaAlaGlnArgThrHis......ArgArg.ProProGlyCy 646
         |||||||||
377 oProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProP 394

22 CCCA 19
   |||
646 sPro 647

seq_name: SwissProt_40:EXTN_TOBAC

seq_documentation_block:
ID   EXTN_TOBAC   STANDARD;   PRT;   620 AA.
AC   P13983;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN   HRGPNT3.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. XANTHI; TISSUE=Leaf;
RX   MEDLINE=90128263; PubMed=2612909;
RA   Keller B., Lamb C.J.;
RT   "Specific expression of a novel cell wall hydroxyproline-rich
RT   glycoprotein gene in lateral root initiation.";
RL   Genes Dev. 3:1639-1646(1989).
CC   -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC   THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC   MAIN ROOT.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC   SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC   GLYCOSYLATED.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X13885; CAA32090.1; -.
DR   PIR; S06733; S06733.
KW   Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW   Hydroxylation.
FT   SIGNAL   1   20   POTENTIAL.
FT   CHAIN   21   620   EXTENSIN.
FT   REPEAT   70   73   H-A-P-P.
FT   REPEAT   148   151   H-A-P-P.
FT   DOMAIN   229   242   2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT   REPEAT   229   235   1.
FT   REPEAT   236   242   2.
FT   DOMAIN   205   620   CONTAINS THE SER-PRO(4) REPEATS.
FT   DOMAIN   499   600   3 X APPROXIMATE TANDEM REPEATS.
SQ   SEQUENCE   620 AA;   65406 MW; 641DD2278AB28524 CRC64;

alignment_scores:
         Quality: 214.00   Length: 424
         Ratio: 1.251   Gaps: 22
Percent Similarity: 40.330   Percent Identity: 27.358

alignment_block:
US-09-786-136-4 x EXTN_TOBAC

Align seg 1/1 to: EXTN_TOBAC   from: 1 to: 620

269 CTTCAGACAAGTAAGGTCAGCTCGCCCCTCCTGCTTCACTTGCCCTGT 318
    ||||||::::::|| :: :|| |||

865 GGCACTGTCACCCTGCCACACCAGGGCCCAGAAGTGCCCAGGGCAATG 816
    |||||| ||| ::: :::|||||||| |:::
394 roGlyCys..ProArgSerAlaArgAsnProGlyCysProArgThr... 408

815 AGGGCCGAGGTGGGCGTGAGGCTGGCGCGGTGCACGGCAAAGGCCGAGGCA 766
    ||| ||| ||| :::||:||| :||||
409 ....Trp.........ArgArgSerGlyAl 415

765 CGGCAGAGGGCGGCTCCTGCCGAGATGCGCACGCGGCCGCAGAAGGGCGCGTC 716
    ||||| ||| ||| || :::|||| |:::
415 aGlnArgGlyHisProProGlyAlaGlyGlnArgProSerGlyProT 432

715 CAGGGCGAAGGAGCCTGGTGCCTCCTGAGCCCGCACCAAGGGCAGGTAGT 666
    ||||||||| || :::||| ::: :::||||||||| :||||
432 hrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla... 447

665 CATGGTCGCTCGCCTTCACTGTTGTTGGCCGGCCCGGGTCAGGGCTACAG 616
    ||| ||| || ||||||||| |||
665 ProGlyProGlyGlyGlyAlaAl 455

615 AGTTTGTGACGGCGCCATGCTGAGCCGACAGGGAGGACGGGGAGGTG. 567
    ||| ::||||:|::   :::||| :::||| ::
455 aValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaA 472

566 .......GGCGGCGAGGACCTTCCGTGGGAGTGGCAGGGTCCC........ 531
    ::::::: |||||| ::: :::||||||
472 spProProAlaAlaAlaArgLeuProProGluProGlnGluProArgLeu 488

530 CCAAAGCAC..........GCTTGGAGTTGCCATTTGGTCCGAGTCCCAG 490
    |||:::: ||| :::||| ||||| ||||| |||
489 ProGlnAspLeuAlaAlaGlnArgCysProAlaGlyCysProProProTh 505

489 GCGCCTGCTGGGAGAGCCTTCTGAAGCTCTTAGAGTAGACAACAGTGGG....A 511
    | —
505 rArgSerGlyAlaAla............................... 511

439 CAGAGGAGCTGCATCTA......GATGGCTGTGAGAGGCAGGAGGAGGGTTC 396
    ||::::: ||| ||||| |:::
439 laGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsn 527

395 CTGGGGGCTCGGCGGGCTCTCGAGCTGCACACCTCCAGGGCCCCAGGCAGG 346
    |||::: :::::   ::::: :::::
528 ProGlyCysProArgThrTrpArgArgSerGlyAlaGlnArgArgGlyHi 544

345 ACCTCCCTAGTAGTAGCATCCACAGCTACAGGGCAAGTGAAGCAGGAGGG 296
    ||| ||| ::::: ||| ||| |||
544 sPro.ProProGlyAla......GlyGlnArgProSerGlyProThrGly 558

295 GGCAGAGCTGACCTTACTTGTCTGAAGGGCTCGAAGCCAGCACCTTCCGGG 246
    |||||| :::::: |||||
559 GlyArgProAlaAlaProGlyAlaProGlyThr...ProAlaAlaProGl 574

245 CATGGCCGTGTGGAACCAAGGAGCAGTCTGTGGAGCCCCACCACCCACCC 196
    | ::: |||:::   :::||| |||
574 yProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 591

195 TCTACAGAGGGCTCATAGAGAAAAAATTGCAGTGTTCTCCCCAGCTTCTCCC 146
    |||||||:::|||
591 rgGlySerGly...................ProAlaAspPro 598

145 TCTATCGACTTTAACAGAGTAGATTGCCTTGCTAGGCCCAGAACCAGCTC 96
    :::   :::||||| |||||
599 ProAlaAlaAlaArgLeuProProGluProGlnGluProArgLeuProGl 615

95 AGAC.......................TCCCCAACTTGACAAGAA 73
    || ::||| ||||||
615 nAspLeuAlaAlaAlaAlaGlnArgCysProAlaGlyLeuProProThrArgs 632

72 GTGATCAGGCTGAATCTTCAACCCATAGACAGAGACAAGAGCCCCTGGCTG 23
   || ::: |||| ||||| ||| ||| || |||||||||||

```
1121 GACCTGGCGCCCTCAAGGCCCCAGCCCGAGGGCTCTTCCTGCTACTGCG 1170
         |||   ||| |||||||||||      ||||||||
 247 erProLeuArgHisLeuProProSerProArgArgGlnProGlnProPro 263

1171 CCACTGGGACCAAAACCTCAGCTACACCTGCTGTGTCACAGCCCAGCGA 1220
         ::::||| |||  |||                  ||||||||
 264 ThrTrpTyrProProAlaTyrAlaGlnSerProGlnProSerPr 280

1221 ACGTGTGAAGGCTGCCCCTCTGCTTGGGCTGGCGCCCCACCCAACACA 1270
         ||||||||      :::  ||||      :||||||
 280 oThrTyrSerPro.ProProProThrTyrSerProProSerPro 296

1271 CTCAAGTCACTGCGCCCAGGCTGGCCTCTTGGTCTGTGGGAAAGTGTAG 1320
         ::: |||  ||||| |||    |||
 297 IleTyrSerProProProAla.................... 304

1321 GCTGGTGCCAGCCTGTCCCCACTGCTTCTTACTCCCCTGGAGCCCT 1370
         |||||              :::|:::||:|||
 305 ...TyrSerProSerProProThrProThrProThrPheSerProP 320

1371 CTTGCCCCACAAAAGTGCCTGCCTGTGCTCTCCCCTCCTCCTCCACC 1420
         ||| ::: ||||| ||||||||
 320 roProProAlaTyrSerProProThrTyrSerProProProThr 336

1421 CCACTCACACTCCCCTCC 1438
         ||| ||||||||
 337 TyrLeuProSer 342


seq_name: SwissProt_40:NME4_MOUSE

seq_documentation_block:
ID   NME4_MOUSE     STANDARD;      PRT;   1323 AA.
AC   Q03391;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
DE   D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D).
GN   GRIN2D.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93050214; PubMed=1385220;
RA   Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA   Inoue Y., Mishina M.;
RT   "Cloning and expression of the epsilon 4 subunit of the NMDA receptor
RT   channel."
RL   FEBS Lett. 313:34-38(1992).
RN   [2]
RP   REVISIONS.
RA   Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA   Inoue Y., Mishina M.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC       POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC       SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC   -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   --------------------------------------------------------------------------
```

```
  13 LeuGlnSerThrAlaAlaLeuLeuSerLeu.....................Va 22

 319 AGCTGTGGATGCTACTACTAGG.............................G 341
         |||:::|||||||||::
  22 lAlaAlaGluAlaThrThrGlnTyrGlyGlyTyrLeuProProProValT 39

 342 AGTCCTGCCTGGGCCCTCAGAGGTGTGCAGCTGAGCCCGCTGAGCCC 391
         |||::::|| |||         ||| |||:::::::
  39 hrSerGlnProProProSerIleGlyLeuSerProSerProSerAlaPro 55

 392 CCAGGAACCCCTCCTGCCTCTCCAGGCCATCTAGATGCAGCTCCTCTGCC 441
         ||||||||   |||:::::||:|    ||| |||
  56 ThrThrThrPro..SerArgGlyHisVal.......ProSerPr 68

 442 CACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTTCCAGCAGGCGCCT 491
         |                  :::::: |||   |||
  68 o...................ArgHisAlaProArgHisAlaT 77

 492 GGGACTCGGACCAAAAATGGCAACTCCAAGCGTGCTTTGGGGGACCCTGCC 541
         :::::::|||::::        ||||||||||
  77 yrProProSerHisGlyHisLeuProSerValGlyGlyPro... 92

 542 ACTCCCACGGAAGGT.......CCTCGCCC.........CCACCTCC 573
         ||:::||| |||   |||||||| |||
  93 ProHisArgGlyHisLeuProProSerArgGlyPheAsnProProPr 109

 574 CCGTCCTCC.CTGTCGGCTCAGCATGGGCCGCGCCGTCACAAACTCTGTAGC 622
         | ||| :::||| ||:|||| ||||||||
 109 oSerProValIleSerProSerHisProProProSerTyrGlyAla...P 125

 623 CCTGACCC.......GGGCCAGGCCAA..CAACAGTGAAGGCAGCGACCA 663
         ||||||| ||||||||
 125 roProProSerHisGlyProGlyHisLeuProSerHisGlyGlnArgPro 141

 664 TGACTACCTGCCCTTGGTGCGGCT.......GCAGGAGGCAC........ 698
         |||::: ||| |||     ::||||
 142 ProSerProSerHisGlyHisAlaProSerGlyHisHisThrProPr 158

 699 .......CAGGCTCCTTCCGCCTGACCGCGCCCTTCCGCGCCCGTGTGCG 741
         ||| |||  :::      ||||||||| |||
 158 oArgGlyGlnHisProSerHisArgArgProSerPro........ 172

 742 CATCTGCCAGGAGC..........GCCTCTGCCGTGCCT........... 770
         ||||||||           :::|||  |||
 173 ..SerArgHisGlyHisProProProThrTyrAlaGlnProProPro 188

 771 CGCCCTTTGCCGTGCCACCCGCGGCCCCAGCCCACCTCGGCCTCATTG 820
         |||::::  :::|||          ||||||
 189 ThrProIleTyrSerProSerProGlnValGlnProProProThrTyrSe 205

 821 CCCTGGCCACTTCTGGGCCTGTGTTGGCCAGGTGACAGTGCCACCGC 870
         |||           |||          |||
 205 rPro..............................ProProP 209

 871 CTCCTGCAGCCCGCTCCCCAGCTCGGGCTCTCTGAGGGCCAGGCCAGTGG 920
         |||::::|||
 209 roThrHisValGlnPro............................ 214

 921 ACAGTGGCGCGGGGCTCAGACACCGAGGGCTCCGATGGGCCGGAAGGGCTG 970
         ||||||  |||||     |||
 215 ..............ThrProSerProSer.ArgGly........ 222

 971 GGCGGCACCGACCTGCGGGGCCGACCTGGGCCACCTGCGTAGCACTCGC 1020
         |||::::|| |||        :::::::
 223 .....HisGlnProGlnProProThrHisArgHisAlaProProThrHis 237

1021 CTGGCTGGAGCACCGATGCGCCGCTGCCTTCGACGAGTGGGAACTGACAG 1070
         :::  ||  |||
 238 ArgHisAlaProThrHis.............................. 244

1071 CGGCCAAGGCTGATTGCTGGCGTGCGGGCCCCCAGCACTTGCCTGACGGCCTT 1120
         :::|||
 245 .............................GlnProS 247
```

```
DR   EMBL; D12822; BAA02254.1; -.
DR   PIR; S27224; S27224.
DR   HSSP; P19491; 1GR2.
DR   MGD; MGI:95823; Grin2d.
DR   InterPro; IPR001320; Ion_glut_receptor.
DR   InterPro; IPR001508; NMDA_receptor.
DR   InterPro; IPR001311; SBP_glut_receptor.
DR   Pfam; PF00060; lig_chan; 1.
DR   PRINTS; PR00177; NMDARECEPTOR.
DR   SMART; SM00079; PBPe; 1.
KW   Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW   Ionic channel; Magnesium.
FT   SIGNAL     1      27       POTENTIAL.
FT   CHAIN      28     1323     GLUTAMATE [NMDA] RECEPTOR SUBUNIT
                                EPSILON 4.
FT   DOMAIN     28     581      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   582    601      1 (POTENTIAL).
FT   DOMAIN     602    623      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   624    643      2 (POTENTIAL).
FT   DOMAIN     644    653      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   654    672      3 (POTENTIAL).
FT   DOMAIN     673    841      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   842    862      4 (POTENTIAL).
FT   DOMAIN     863    1323     EXTRACELLULAR (POTENTIAL).
FT   DOMAIN     278    283      POLY-GLY.
FT   DOMAIN     905    913      POLY-PRO.
FT   DOMAIN     1030   1035     POLY-ALA.
FT   DOMAIN     1197   1201     POLY-PRO.
FT   SITE       639             FUNCTIONAL DETERMINANT OF NMDA
                                RECEPTORS (BY SIMILARITY).
FT   CARBOHYD   89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   349    349      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   363    363      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   464    464      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   566    566      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1323 AA;  142907 MW;  8AE9878F90DD0921 CRC64;

alignment_scores:
      Quality:        211.50          Length:       443
      Ratio:          1.347           Gaps:          23
Percent Similarity:   35.440     Percent Identity:  25.056

alignment_block:
us-09-786-136-4 x NME4_MOUSE    ..

ign seg 1/1  to: NME4_MOUSE   from: 1   to: 1323

 353 GGGGCCTGCAGGTGCAGCTCAGAGCCCGCTGAGCCCCCAGGAACCCC 402
     |||  ||||||:|| |||||||::|||| || |||||:||:|| ||
 889 GlyMetTyrSerCysSerAlaGluAlaAlaProProProAlaLysPr 905

 403 TCCTGCCTCTCCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGTGTCT 452
     ||||                   |||||||||||||::: :::|
 905 oProPro.................ProProGlnProLeuProSerProAlaT 917

 453 AC.................................. 454
     ||
 917 yrProAlaAlaArgProProGlyProAlaProPheValProArgGlu 933

 455 ..................TCTAAAGGACTTCAGAGAGGCTC 477
                       ::||||||   ||
 934 ArgAlaAlaAlaAspArgTrpArgArgAlaLysGly.....ThrGlyPr 948

 478 TCCAGCAGGCGCCTGGGACTTCGGAC.................. 502
     |||::|||||||  :::|||
 948 oProGlyAlaAlaLeuAlaAspGlyPheHisArgTyrTyrGlyProI 965

 503 .............................CAAAATGGCAACTCCAAGCGT 523
                                  :::: :||||  :::
 965 leGluProGlnGlyLeuGlyLeuGluGluAlaArgAlaAlaProArgGly 981

 524 GCTTTGGGGACCCTGCCACTCCC.....ACGGAAGGTCCTCGCCGCCC 567
```

```
 982 AlaAlaGlyArgProLeuSerProProThrThrGlnProProGlnLysPr 998
     |||  |||          :::||| |||   :::::::|:||
 568 ACCTCCCGCTCCTCCCTGTCGGCTCAGCATGGCCGCCGTCACAAACTCT 617

 998 oProPro.......................................... 1000
 618 GTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGAC 667

1001 ...............................................Ser 1001
 668 TACCTGCCCTGGTGCGGCTGCAGGAGGCA..................... 697
     |||:::  :::||||| ||||||
1002 TyrPheAlaIleValArgGluGlnGluProAlaGluProProAlaGlyVal 1018

 698 ....CCAGGCTCCTTCCGCCTGACGCGCCCCTTCTGGCCGCCGTGCGCCA 743
         ||||||||
1018 aPheProGlyPheProSerProProAlaProAlaAlaAlaAlaAlaAlaA 1035

 744 TCTCGCAGGAGGCCTCTGCCGT...............GCCTGCCCCTTT 778
     |||||||
1035 laValGlyProProLeuCysArgLeuAlaPheGluAspGluSerProPro 1051

 779 GCCGTGCACCGCGCCAGCCTCAGCCCACCTCGGCCTCATTGCCCTGGGC 828
     |||   |||      |||
1052 AlaProSerArg.........TrpProArgSerAspProGluSerGlnPr 1065

 829 ACTTCTGGCCCTGTGTTGGCCAGGGTGACAGTGCCACGGCCTCCTGCA 878
     |||:: |||:::||  ||||::|||
1065 oLeuLeuGlyGlyGlyAlaGly...GlyProSerAlaGlyAlaProThrA 1081

 879 GCCCGTCCCCAGCTGGGCTCTGAGGGGCCAGCCAGTGGACAGTGGG 928
     ::||| |||           |||   |||
1081 laProProProArgThrAlaProProCysAlaTyrLeuAspLeu 1097

 929 CGGGGCTCAGACACCGAGGGCCTCCGATGGGCGGAAGGGCTGGGCGGCAC 978
     :::  |||  ::||||:::||||| ||||:::|||||||||
1098 GluProSerProSerAspSerGluAsp...SerGluSerLeuGlyGlyAl 1113

 979 CGACCTGCGGGGCCGGGACCTGGGCACCTGGGCCAGCCACTCGCCTGGCTGG 1028
     ::: ||| ||            |||                |||
1113 aSerLeuGlyLeuGlu...............ProTrp.Trp 1122

1029 AGCACCGATGCGCGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAG 1078
     ::: || ||| |||  :::::
1123 PheAlaAspPheProTyrProTyrAlaGlu.............Ar 1133

1079 GCTGATTGCTGCTCGCGGGCCCAGCACTTCCTGACGGGCCTTGACCTGGC 1128
     ||||          |||||   |||
1133 gLeu..............GlyPro...................P 1137

1129 CGCCCTCAAGG................................. 1139
     |||||
1137 roProGlyArgTyrTrpSerValAspLysLeuGlyTrpArgAlaGly 1153

1140 ............CCGCAGCCCGAGGGCTCTTCCTGCTACTGCGCCAC 1174
                 |||   ||| :::|||||
1154 SerTrpAspTyrLeuProProArgGlyGlyProAlaTrpHisCys...... 1168

1175 TGGGACCAAAACCTGCAGCTACACCTGCTGTCTACAGCCCAGCGAACGT 1224
                         |||
1169 .....................ArgHisC 1171

1225 GTGAAGGCTGCCCCCTGCTGTGGCGTGGCGCCCACCCAACACACTCA 1274
     |||||           |||
1171 ysAlaSerLeuGluLeuLeu..........ProProProArgHisLeu 1183

1275 AGTCACTGCCGCCCAGGGCTGGCCTCTTGGTCGTGCTGGGAAAGTGTAGGCTG 1324
     :: :: |||| |||||
1184 SerCysSerHisAspGlyLeuAspGly..............GlyTyr 1194

1325 GTGGCCAGCCTGTCCCCCACTGCTTCTTACTCCTCCTGGAGC......C 1368
     |  |||||||||::::||||:::
```

```
1194 PTrpAlaProPro...............ProProTrpAlaAlaGlyP 1206
1369 CTCTTGCCCCACAAAAAGTGCCTGCCTGTGCTCTCCCTCCTCCTCCCA 1418
     ||  ||||||  :::::|| |||  |||
1206 roProAlaProArgArgAlaArgCysGlyCysPro..........Arg 1218
1419 CCCCACTCACACTCCCCTCCCATCCTCT 1445
     ||||||  |||  |||  ||  :::|||
1219 ProHisProHisArgProArgAlaSer 1227

seq_name: SwissProt_40:NME4_RAT

seq_documentation_block:
ID   NME4_RAT       STANDARD;        PRT;   1323 AA.
AC   Q62645; Q63381; Q63382; Q63729; Q63730;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
DE   D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D).
DE   GRIN2D.
OX   Rattus norvegicus (Rat).
OX   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 1265-1356 FROM N.A.
RP   (ISOFORM 2).
RC   STRAIN=Sprague-Dawley; TISSUE=Forebrain;
RX   MEDLINE=93155102; PubMed=8828958;
RA   Ishii T., Moriyoshi K., Sugihara H., Sakurada K., Kadotani H.,
RA   Yokoi M., Nkazawa C., Shigemoto R., Mizuno N., Masu M., Nakanishi S.;
RT   "Molecular characterization of the family of the N-methyl-D-aspartate
RT   receptor subunits.";
RL   J. Biol. Chem. 268:2836-2843(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=94206533; PubMed=7512349;
RA   Monyer H., Burnashev N., Laurie D.J., Sakmann B., Seeburg P.H.;
RT   "Developmental and regional expression in the rat brain and functional
RT   properties of four NMDA receptors.";
RL   Neuron 12:529-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RA   Boulter J., Pecht G.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC       WITH HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC       SENSITIVITY TO MAGNESIUM. MEDIATED BY GLYCINE.
CC   -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 AND 2 (SHOWN HERE); MAY BE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, MAINLY IN THE SUBCORTICAL
CC       REGION.
CC   -!- DEVELOPMENTAL STAGE: Already detected in embryonic stages, peaks
CC       at postnatal day 7, and decreases thereafter to adult levels.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D13213; BAA02500.1; -.
DR   EMBL; D13214; BAA02501.1; -.
DR   EMBL; L31611; AAC37646.1; -.
DR   EMBL; L31612; AAC37647.1; -.
DR   EMBL; U08260; AAA17833.1; -.
DR   HSSP; P19491; 1GR2.
DR   InterPro; IPR001320; Ion_glut_receptor.
DR   InterPro; IPR001508; NMDA_receptor.
DR   InterPro; IPR001311; SBP_glut_receptor.
DR   Pfam; PF00060; lig_chan; 1.
DR   PRINTS; PR00177; NMDARECEPTOR.
DR   SMART; SM00079; PBPe; 1.
KW   Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW   Glycoprotein; Ionic channel; Magnesium; Alternative splicing.
FT   SIGNAL          1     27       POTENTIAL.
FT   CHAIN          28   1323       GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT                                  EPSILON 4.
FT   DOMAIN         28    580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      581    601       1 (POTENTIAL).
FT   DOMAIN        602    619       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      620    638       2 (POTENTIAL).
FT   DOMAIN        639    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      655    679       3 (POTENTIAL).
FT   DOMAIN        680    843       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      844    862       4 (POTENTIAL).
FT   DOMAIN        863   1323       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN        278    283       POLY-GLY.
FT   DOMAIN        905    913       POLY-PRO.
FT   DOMAIN       1030   1035       POLY-ALA.
FT   DOMAIN       1197   1201       POLY-PRO.
FT   SITE          639    639       FUNCTIONAL DETERMINANT OF NMDA
FT                                  RECEPTORS (BY SIMILARITY).
FT   CARBOHYD       89     89       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      349    349       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      363    363       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      464    464       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      566    566       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      639    639       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC     1265   1323       CPRAAPTRRLTGPSRHARRCPHAAHWGPPLPTASHRRHRGG
FT                                  DLGTRRGSAHFSSLESEV -> RPCPHRTGDTGAGTWAHA
FT                                  GALRISPAWSPRYDAAPAPTPTPAAPSVSAGHGPRGRAKWT
FT                                  GPSWVGKDRNGPGRTPPGAASCAPTPFALGEL (IN
FT                                  ISOFORM 1).
FT   CONFLICT       25     25       A -> V (IN REF. 3).
FT   CONFLICT       47     47       P -> Q (IN REF. 3).
FT   CONFLICT       67     67       G -> V (IN REF. 2; AAC37646).
FT   CONFLICT       94     94       R -> P (IN REF. 2).
FT   CONFLICT      305    305       R -> A (IN REF. 2).
FT   CONFLICT      635    635       A -> G (IN REF. 3).
FT   CONFLICT      974    974       E -> D (IN REF. 3).
FT   CONFLICT     1253   1253       A -> G (IN REF. 3).
FT   CONFLICT     1266   1267       PR -> TT (IN REF. 3).
SQ   SEQUENCE   1323 AA;  143100 MW;  40F7D60192579564 CRC64;

alignment_scores:
       Quality:    210.50            Length:     441
       Ratio:        1.341           Gaps:        22
Percent Similarity:  35.601     Percent Identity:  24.717

alignment_block:
US-09-786-136-4  x NME4_RAT   ..

Align seg 1/1  to: NME4_RAT  from: 1  to: 1323

353 GGGGCCTCGCAGGTGTCGCAGCTCAGAGCCCGTCAGCCCGAGCCCCAGGAACCCC 402
    |||          :::  |||||||:::||  ||||||:::|| ||
889 GlyMetTyrSerCysCysSerAlaGluAlaAlaProProAlaLysPr 905

403 TCCTGCCTCCACAGCCATCTAGATGCAGTCTAGATGAGCTCCTCTGCCCACTGTGTCT 452
    ||||                         |||||||||:  :::|
905 oProPro...................ProProGlnProSerProAlaT 917

453 AC.......................................... 454
    ||
917 yrProAlaAlaArgProProGlyProAlaProPheValProArgGlu 933
```

Left column:

```
455  ..................................TCTAAAGGACTTCAGAGAGGCTC  477
                                         ::::||||||  |||
934  ArgAlaAlaAspArgTrpArgArgAlaLysGly.....ThrGlyPr  948

478  TCCAGGCGCCGCCTGGGACTCGGAC..............................  502
     |||::||||||    :::|||
948  oProGlyAlaAlaIleAlaAspGlyPheHisArgTyrTyrGlyProI  965

503  ..................................CAAAATGGCAACTCCAAGCGT  523
                                         :|:|| |||   :::
965  leGluProGlnGlyLeuGlyLeuGluGluAlaArgAlaAlaProArgGly  981

524  GCTTTGGGGACCCTGCCACTCCC.....ACGGAAGGTCCTGCCGCCC  567
     ||| ||| ||| |||  |||   :::||:||
982  AlaAlaGlyArgProLeuSerProProThrThrGlnProProGlnLysPr  998

568  ACCTCCCCGTCCTCCGTCGGCTCAGCATGGGCGGCCGCGCTCACAAACTCT  617
     ||||||
998  oProPro..........................................  1000

618  GTAGCCCTGACCCGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGAC  667
                                                    :::
1001 ................................................Ser  1001

668  TACCTGCCCTTGGTGCGGCTGCAGGAGCA.....................  697
     |||:: ||:||||| ||||||
1002 TyrPheAlaIleValArgGlnGluProThrGluProProAlaGlyAl  1018

698  ..CCAGGCTCCTTCCGCCTGACGCGCCCTTCTGCGCGCGTGTGCCGA  743
     ||||||                |||||:::||||||||||:::
1018 aPheProGlyPheProSerProAlaProProAlaAlaAlaAlaAlaA  1035

744  TCTCCGAGAGGCCCTCGCCGT....................GCCTCGCCCTTT  778
     ||||||||                        ||||||||
1035 laValGlyProProLeuCysArgLeuAlaPheGluAspGluSerProPro  1051

779  GCCGTGCACCGGCCAGCCTCAGCCCACCTCGGCTCATTGCCCTGGGC  828
     ||| |||  ||| :::||
1052 AlaProSerArg.........TrpProArgSerArgSerProGluSerGlnPr  1065

829  ACTTCTGGGCCTGCTGTTGGCCAGGGTGACAGTGCCACGGCCTTCCTGCA  878
     :::  ||||||||| |||:::|| |||||| |||
1065 oLeuLeuGlyValGlyAlaGly...GlyProSerAlaGlyAlaProThrA  1081

879  GCCCGTCCCCAGCTCGGGCTCTGAGGGGCCAGGCCAGGTGGACAGTGGG  928
     ::||| |||             :::
1081 laProProProArgProArgAlaAlaProProProCysAlaTyrLeuAspLeu  1097

929  CGGGGCTCAGACACACCGAGGCCTCCGATGGGCGAAGGCTGGGCGGCAC  978
     :::  ||| :::::::::|||||::   ||||||:::||||:
1098 GluProSerProSerAspSerGluAsp...SerGluSerLeuGlyGlyAl  1113

979  CGACCTGCGGGGCCGGACCTGGGCCACTGGGCTAGCACTCGCCTGGCTGG  1028
                                         ||| |||
1113 aSerLeuGlyLeuGluGlu...............ProTrp.Trp  1122

1029 AGCACCGATGCGCCCGCTGCCTTCGACGAGTGGGAACTGACAGCCGCCAAG  1078
     :|||| |||| |||  :::::                 ||
1123 PheAlaAspPheProTyrProTyrAlaGlu........................Ar  1133

1079 GCTGATTGCTGGCTGCGGGCCCAGCACTTGCCTGACGCGGCCTTGACCTGGC  1128
     ||||                          ||||||
1133 gLeu..........GlyPro......................P  1137

1129 CGCCCTCAAGG..................................  1139
     |||||::|||
1137 roProGlyArgTyrTrpProSerValAspLysLeuGlyGlyLeuTyrProArgAlaGly  1153

1140 ..........CCGCAGCCCGGGCTCTTCCTGCTACTGCGCCAC  1174
               ||| ||| |||::|||
1154 SerTrpAspTyrLeuProProArgGlyGlyProAlaTrpHisCys......  1168

1175 TGGGACCAAAACCTGCAGCTACACCTGCTGTGCTACAGCGCCAGCGAACGT  1224
```

Right column:

```
                                         |||
1169 ................................ArgHisC  1171

1225 GTGAAGGCTGCCCCTGCTGCTGGGCTGGCGCCCACCCACCACCACTCA  1274
     || :::        ||||||   |||||:::||
1171 ysAlaSerLeuGluLeuLeu..............ProProProArgHisLeu  1183

1275 AGTCACTGCGCGCCCCAGGCGCCGCCTCCTGGTGCTGGGAAAGTGTAGGCTG  1324
     |||   ::: ||||| :::||                |||||
1184 SerCysSerHisAspGlyLeuGlyLeuAspGly.........GlyTyr  1194

1325 GTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCGGAGCCCTCTTG  1374
     |||||:||| ||| ||                   ||||||||||:::
1194 pTrpAlaProProPro................ProProTrpAlaAlaGlyP  1206

1375 CCCCACAAAAGTGCCTGCCTGCTCTCCTCCTCCCTCCTCCCCCCAC  1424
     ||| |||      ||||::|||:::|||    |||||
1206 roProProArgArgAlaArgAlaCysGlyCysPro......ArgProHis  1220

1425 TCACACTCCCTCCATCCTCT  1445
     ||| |||    :::|||
1221 ProHisArgProArgAlaSer  1227

seq_name: SwissProt_40:EBN1_EBV

seq_documentation_block:
ID  EBN1_EBV    STANDARD;    PRT;   641 AA.
AC  P03211;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  EBNA-1 nuclear protein.
GN  BKRF1.
OS  Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC  Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC  Gammaherpesvirinae; Lymphocryptovirus.
OX  NCBI_TaxID=10377;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=84270667; PubMed=6087149;
RA  Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA  Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA  Tuffnell P.S., Barrell B.G.;
RT  "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL  Nature 310:207-211(1984).
RN  [2]
RP  SEQUENCE OF 1-26 FROM N.A.
RX  MEDLINE=86259739; PubMed=3460083;
RA  Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT  "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT  proteins: a probable transcriptional initiation site.";
RL  Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN  [3]
RP  SUBCELLULAR LOCATION.
RX  MEDLINE=9026473; PubMed=2161150;
RA  Petti L., Sample C., Kieff E.;
RT  "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT  latent infection nuclear proteins.";
RL  Virology 176:563-574(1990).
RN  [4]
RP  X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX  MEDLINE=96006523; PubMed=7553871;
RA  Bochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA  Edwards A.M., Frappier L.;
RT  "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT  virus origin-binding protein EBNA 1.";
RL  Cell 83:39-46(1995).
CC  -!- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC      MAINTENANCE REPLICATION OF EBV EPISOME. TRANSACTIVATING FACTOR
CC      FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
CC      ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC      WITH THE NUCLEAR MATRIX.
```

```
DR   EMBL; V01555; CAA24816.1; -.
DR   EMBL; M13941; AAA45889.1; -.
DR   PIR; A03773; QQBE31.
DR   PIR; S33021; S33021.
DR   PDB; 1VHI; 23-DEC-96.
DR   TRANSFAC; T00211; -.
DR   InterPro; IPR004186; EBNA1.
DR   Pfam; PF02905; EBNA1; 1.
KW   Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW   3D-structure.
FT   DOMAIN       87    352       GLY/ALA-RICH.
SEQUENCE   641 AA;  56427 MW;  4D1G1653E16FC341 CRC64;

alignment_scores:
          Quality:  205.00          Length:    474
          Ratio:    0.919           Gaps:       21
Percent Similarity:  47.046   Percent Identity:  25.738

alignment_block:
US-09-786-136-4/rev x EBN1_EBV  ..

Algn seg 1/1  to: EBN1_EBV  from: 1 to: 641

1434 GGGAGGTGAGTGGGGTGGGAGGAGGAGGAGAGCACAGGCAGGCACTT 1385
     ||:|:::|||:::||||:||||  ||  :::  ||  |||
 161 GlyGlyAlaGlyGlyAlaGlyAlaGlyAlaGlyGlyAlaGlyAla..... 175

1384 TTTGTGGGGCAAGAGGGCTCCAGGGAGGAGTAAGCAGTGGGGGAC 1335
     || |||||| ||:::||||  ::||||||||:::||||||
 176 ......GlyAlaGlyGlyAlaGlyAlaGlyAlaGlyAlaGlyGlyAlaGlyA 190

1334 AGGCTGGCACCAGCCTACACTTTCCCAGCACCAAGAGGCCAGCCCTGGGC 1285
     ::||||||||:::|||
 190 laGlyGlyAlaGly.........GlyAlaGlyAlaGlyGly 200

1284 GGCAGTGACTTGAGTGTGTTGGGTGGGCGGCCAAGCCAGCAGGGGG 1235
     ||||||| ||:::||||  :::|||||:::||||||
 201 GlyAlaGlyAlaGlyGlyAlaGlyAlaGlyAlaGlyAlaGlyAlaGlyGl 217

1234 CAGCC.................TTCACACGTTCGCTGGGCTGTAGC 1206
     | ||||                :::|||:::|||
 217 y.AlaGlyAlaGlyAlaGlyGlyAlaGlyAlaGlyAlaGlyAlaGly 233

1205 ACAGCAGGTGTAGCTCGAGGTTTGGTCCCAGTGGGCAGTAGCAGGAAG 1156
     |||||| |||:::||| ||:::||||||||  |||||
 234 GlyAlaGly..AlaGlyGlyAlaGlyAlaGlyGlyAlaGlyAlaGlyGl 249

1155 AGCCCTCGGGCTGCGGGCCTTGAGGGCGCCAGGCCGTCAGGGCAA 1106
     |||  ::|||||||:::  || |||  :::
 249 yAlaGlyAlaGlyGlyAlaGlyAlaGlyGlyAlaGlyAlaGlyGlyA 266

1105 GTGCTGGGCCCCGCCCCCAGCCAGCAATCAGCCTTGGCCCTGTCAGTTCCCACT 1056
     :::|||| ::|||||||||:::  ||| :::  |||:::
 266 laGlyAlaGlyAlaGlyGly.GlyAlaGlyAlaGly......... 279

1055 CGTCGAAGGCAGCGGCGCCATCGGTGCTCCAGCCAGGCGAGTGCTACGGCA 1006
     :::||||||||||||||:::  ||  :::  |||
 280 ...AlaGlyGlyAlaGlyGlyAlaGlyAlaGlyGlyAlaGlyGlyAlaGlyA 295

1005 GTGGCCCAGGTCCGGCCCCCGAGGTCGGTGCCGCCCCCAGCCCTTCGGCCCC 956
     :::||| ||| |||:::||:::|||||||:::
 295 laGlyAlaGlyGlyAlaGlyGlyAlaGlyAlaGlyAlaGly.......... 307

 955 ATCGGAGGCCTCGGTGTCTGAGCCCGCCCACTGTCCACCTGGCCTGGCC 906
     ||||||  ||:::
 308 AlaGlyGlyAlaGlyAla.................................. 313

 905 CCTCAGAGCCCGAGCTGGGGACCGGGCTGCAGGAGGCCGTGGCCACTGTCA 856
     :::|||||| ::||||||||| ::|||||||
 314 ........GlyGlyAlaGlyAlaGlyAlaGlyAlaGlyAlaGlyAla..... 324

 855 CCCTGGCCAACACCAGGGCCCAGAAGTGCCCAGGGCAATGAGGGCCGAGGT 806
     ||| |||
 325 ..........................................GlyGlyGly 327

 805 GGGGCTCAGGCTGGCGCGGTGCACGGCAAAGGGCGGAGGCACGGCAGAGGC 756
     ||| ||:::  ||:::||  ||||||||| ::: ::
 328 GlyArgGlyArgGlySerGlyGlySerGlyArgGlyArgGlySerGlyGl 344

 755 GCTCCTGCGAGATGGCCACACGCCAGAAGGGGCGCGTCCAGGCCGAAG 706
     ||| :: ::|||  :::||||| ::||||| :::|||||
 344 yArgGlyArgGlySerGlyGlyArgArgGlyArg.GlyArgGluArg 360

 705 GAGCCTGGTGCCTCCTGCAGCCGCCACCAAGGGCAGGTAGTCATGGTCGCT 656
     ||:::||| :: :::||| :::  ||||||:::|||
 361 AlaAlaArgGlySerArgGluArgAlaArgGlyArg.............. 372

 655 GCCTTCACTGTTGTTGGCCTGGCCCGGGTCAGGGCTACAGAGTTTGTGAC 606
     ||| ||| :: :::||| ||||||:::
 373 .............GlyArgGlyArgGlyGluLysArgP 381

 605 GGCGGCCCATGCTGAGCCCGACAGGGAGGACGGGGAGGTGGGCGGCGAGGA 556
     ||| :::  ||| ::||||||:::|||||
 381 roArgSerProSerGlnSerSerGlySerProProArgArg 397

 555 CCT.....................TCCGTGGGAGTGGCA......... 538
     |||
 398 ProProProGlyArgArgProPhePheHisProValGlyLeuAlaAspTy 414

 537 .................GGGTCCCCAAAGCACCGCTTGGAGTTGCCAT 507
                      :::|||||:::||:::
 414 rPheGluTyrHisGlnGluGlyProAspGlyGluProAspValProP 431

 506 TTTGGTCCGAGTCCCAGGCCGCCTGCTGGAGAGCCTCTCTGAAGTCCTTTA 457
     :: :::  ::|||||||||:::||  :::  ||||
 431 roGlyAlaIleGluGlnGlyLysProAlaAspProGlyGluGlyPro... 446

 456 GAGTAGCAACACAGTGGGCAGAGGAGCTGCATCTAGATGGCTGTGAGAGGC 407
     ::||| |||||:::  || :::|||||
 447 .....SerThrGlyProArgGlyGlnGlyAsp.GlyGlyArgArgLysL 461

 406 AGGAGGGGTTCCTGGGGCTCAGCGGGCTCGAGCTACCACCTGCAGGG 357
     :: |||  :::||||
 461 ysGlyGlyTyrTrpPheGlyLysHisArgGly................. 470

 356 CCCCAGGCAGGACCTCCCTAGTAGTAGCAATCCACAGCTACAGGGGCAAGTG 307
     ::|||| ||:::  :::|||
 471 ...GlnGlyGlySer..........AsnProLysPheGluAsn...... 480

 306 AAGCAGGAGGGGCAGCAGCTGCACCTTACTTGTCTGAAGGGCTCGAAGCC 258
     :::|||||||||:::||:::
 481 ........IleAlaGluGlyLeuArgAlaL 488

 257 .......................AGCACCTCCGGGCATGCCGCTGTGGAAC 231
     ||| |||
 488 euLeuAlaArgSerHisValGluArgThrArgAspGlyThrTrpPheVal 504

 230 CAAGGAGCAGTCGTGTGGAGCCCCACCACCCTTACAGAGGGCTCA 181
     :: ||| :::  ||:::
 505 AlaGlyValPheValTyrGlySerLysThrSerLeuTyrAsnLeuAr 521

 180 TAGAAAAAATTGCAGTGTTCTCCCCAGCTTCTCCCCTCATCGACTTTAAC 131
     :: ||| :::  ||
 521 gArgGlyThrAlaLeuAlaIleProGlnCysArgLeuThrProLeu..... 536

 130 AGAGTAGATTGCCT 117
```

```
              |||||||||||
537  ..SerArgLeuPro 540

seq_name: SwissProt_40:NME4_HUMAN

seq_documentation_block:
ID   NME4_HUMAN      STANDARD;      PRT;   1336 AA.
AC   O15399;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
DE   D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D) (EB11).
GN   GRIN2D.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98149394; PubMed=9489750;
RA   Hess S.D., Daggett L.P., Deal C., Lu C.-C., Johnson E.C.,
RA   Velicelebi G.;
RT   "Functional characterization of human N-methyl-D-aspartate subtype
RT   1A/2D receptors.";
RL   J. Neurochem. 70:1269-1279(1998).
CC   -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC       WITH HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC       SENSITIVITY TO MAGNESIUM. MEDIATED BY GLYCINE.
CC   -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U77783; AAC15910.1; -.
DR   HSSP; P19491; 1GR2.
DR   MIM; 602717; -.
DR   InterPro; IPR001320; Ion_glut_receptor.
DR   InterPro; IPR001311; SBP_glut_receptor.
DR   Pfam; PF00060; lig_chan; 1.
DR   SMART; SM00079; PBPe; 1.
KW   Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW   Glycoprotein; Ionic channel; Magnesium.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   1336       GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT                                EPSILON 4.
FT   DOMAIN       28    583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    584    604       1 (POTENTIAL).
FT   DOMAIN      605    622       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    623    641       2 (POTENTIAL).
FT   DOMAIN      642    657       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    658    682       3 (POTENTIAL).
FT   DOMAIN      683    846       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    847    865       4 (POTENTIAL).
FT   DOMAIN      866   1336       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      281    286       POLY-GLY.
FT   DOMAIN      908    916       POLY-ALA.
FT   DOMAIN     1035   1040       POLY-PRO.
FT   DOMAIN     1209   1213       POLY-ALA.
FT   DOMAIN     1244   1247       POLY-ALA.
FT   SITE        642    642       FUNCTIONAL DETERMINANT OF NMDA
FT                                RECEPTORS (BY SIMILARITY).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    366    366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    467    467       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    569    569       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    642    642       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1336 AA;  143558 MW;  DECC545F3E416680 CRC64;

alignment_scores:
          Quality:   200.00           Length:     455
            Ratio:     1.258             Gaps:      24
Percent Similarity:  34.945   Percent Identity:  24.615

alignment_block:
US-09-786-136-4 x NME4_HUMAN

Align seg 1/1  to: NME4_HUMAN    from: 1    to: 1336

 353 GGGGCCCTGCAGGTGCGAGCTCAGAGCCCGTGAGCCCCAGGAACCCG  402
     |||   |||||| ||||::::|  ||||||::|||  ||||||:: ||
 892 GlyMetTyrSerCysSerAlaGluAlaAlaProProProAlaLysPr  908

 403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCCACTGTTGTCT  452
     ||||                     |||||||||||::::
 908 OProPro........ProProGlnProLeuProSer............  917

 453 ACTCTAAGGACTTCAGAGAGGCTCTCCAGCAGCCGCCTGGGACTCGGAC  502
 917 ..................................................  917

 503 CAAAATGGCAACTCCAAGCGTGCTTGGGGAGACCTGCCACTCCCACGGA  552
                                          |||||  |||:::
 918 ...........................ProAlaTyrProAlaPr  923

 553 AGGTCCTGCGCGCCACCT...........CCCCGTCCTCCCGTCGGC  590
     ||||||            |||        |||||   :::
 923 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaserV  940

 591 TCAGCATGGGCCGCCGTCACAAACTCTGTAGCCCTGACCCGGGCCAGGCC  640
     |||             ||||| |||||||  :::::||  |||||  |||
 940 alAspArgTrpArgThrLysGlyAlaGlyPro...ProGlyGlyAla  955

 641 AACAACAGTGAAGGCAGCGAGCCATGACTACCTGCCCTTG..........  679
     :::  :::::::|||       ::: |||
 956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnGl  972

 680 .......GTGCGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACGCGC  722
        ::: ||| |||||  ||| ||||| :::|||
 972 yLeuGlyLeuGlyLeuGluAla.ArgAla.....................  982

 723 CCTTCTGCCGCCGTGTGCCCATCTCGCAGGAGGCGCCTCTGCCGTGCCTCG  772
     ||||||  |||                    ||
 983 ....AlaProArgGlyAlaAlaGlyArg......................  990

 773 CCCTTGCCGTGCACCGCCCCCAGCCCACCTGGCCTCATTGCC  822
     |||  ::: ||||||:: |||||::: |||::::
 991 ......ProLeuSerProProAlaGlnProProGlnLysPro....P 1004

 823 CTGGGCACTTCTGGGCCCTGGTGTGGCCAGGGTGACAGTG..........  863
     ||::  ::: ::: ::|||:::||:||
1004 roAlaSerTyrPheAlaIleValArgAspLysGluProPro 1020

 864 ..........CCACGGCTCCTGCAGCGCCGTCCCCCAGC  892
             ||||||||      |||||||||||
1021 AlaGlyAlaPheProGlyPheProSerProProAla......ProProAl 1035

 893 TCGGGCTCTGAGG.....  905
     |  |||
1035 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052

 906 ................GGCCAGGCCAGTGGACA  923
                     ::: ||||||||
1052 spGluSerProProAlaProAlaArgTrpProArgSerAspProGluSer 1068

 924 ............GTGGGCGGGGCTCAGACACCGAGGCCCTCC  952
                 :::  |||||||||
```

```
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- FUNCTION: TEGUMENT PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC       EHV-1 24, EBV BPLF1, HVS-1 64, V2V 22, AND HCMV UL48.
CC   ---------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------
DR   EMBL; V01555; CAA24839.1; -.
DR   PIR; A03747; QQBE8.
DR   PIR; S32993; S32993.
SQ   SEQUENCE   3149 AA;   337954 MW;   3DD0C576587313D8 CRC64;


alignment_scores:
        Quality:  199.50              Length:      415
          Ratio:    1.108               Gaps:       23
   Percent Similarity:  43.373   Percent Identity: 26.988

alignment_block:
US-09-786-136-4 x TEGU_EBV

Align seg 1/1  to: TEGU_EBV   from: 1   to: 3149

483  CAGGCGCTGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGGG  532
     :::|||  |||||||                :::::|||  |||:||
279  GluAlaAlaGlyThrProGlyAlaAspSerSerProProValMet......  293

533  GACCCTGCCACTCCCACGGAAGGTCCTCGCCGCCCACCTCCCGTCCTCC  582
     |||||::: |||:::||||        :::||| ||||||||
294  ....LeuProPheGluArgArgIleIleProTyrAsnLeuArgProLeuP  309

583  CT........GTCGGCTCAGCATGGCCGCCGCCCTCACAAACTCTGTA  620
     ||              |||||||  |||:::      |||   |||
309  roSerArgSerPheThrSerAspSerPheProAlaAlaArgTyrSerPro  325

621  GCCCTGACCCCGGCCCAGGCCCAACACA..GTGAAGGCAGCGACCATGAC  667
     ||| ||||:::| |||:::::  |||||        :::||:||||||
326  AlaLysThrAsnSerProProSerProSerAlaAlaProAlaSe  342

668  TACCTGCCCTTGGTGCGGCGTGCAGAGGCACCAGGCTCCTTCCGCCTGGA  717
     ||||||        |||:::     |||||  ||||:::::
342  rAlaAlaProAlaSerAlaAlaProAlaSerAlaAlaProAlaSer.....  357

718  CGGGCCCTTCTGCGCCGCCGTGCGCATCCGCAGGAGCGCCTCGCCGTG  767
     :::|||||    |||||||||  ||||||        ||||||||||:
358  ......AlaAlaPro......AlaSerAlaAlaProAlaSerAlaAla  369

768  CCTCGGCCCTTGCCGTGCACGGCGCCAGCCTCAGCCCCACCT.........  809
     |||                :::||||||||||||||::|||||
370  Pro........AlaSerAlaAlaProAlaSerSerProProLeuPheIl  383

810  ....CGGCCTCATTGCCCTGGGCACTTCTGGGCCCTGGTGT  846
     :::||||||||:::|||||
383  eProIleProGlyLeuGlyHisThrProGly..................  393

847  TGGCCAGGTGACAGTGCCACGGCCTCCTGCAGCCCGTCCCCCAGCTCGG  896
     ||||||       ||||||  |||||||||||
394  ..............ValProAlaProSerThrProProArg  402

897  GCTCTGAGGGCCAGGCCAGTGACAGTGGCGGGGCTCAGACACCGAG  946
     |||:: |||  |||
403  AlaSerSerGlyAlaAla.................ProGl  410

947  GCCTCCGATGGGCGCGAAGGCTGGGCG..........GCAC  978
```

```
1069  GlnProLeuLeuGlyProGlyAlaGlyGlyAlaGlyGlyThrGlyGlyAl  1085
      :::|||||||||
 953  GATGGGGCCGGAAGGGCTGGGCGGCCACCGACCTCGGGGGCCGGACCTGGGC  1002
      |||  |||  ||| |||:::|||
1085  aGlyGlyGlyAlaGlyAlaAlaProAlaProProCysAlaAlaProProP  1102
      ||| |||  ||| |
1003  CACTGCCG..........................  1010

1102  roCysProTyrLeuAspLeuGluProSerProSerAspSer  1118
      || |||
1011  ...........TAGCACTCGCTGCTGGAGCAC  1033

1119  GluSerLeuGlyLyGlyAlaSerLeuGlyLyLeuAspProTrpTrpPheAl  1135
      :::|||  |||      ::||| |||
1034  CGATGCGCCGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAGGCTGA  1083

1135  aAspPheProTyrProTyrAlaGlu............ArgLeu..  1145
      |||  |||      |||:::::
1084  TTGCTGGCTGGGGCCCAGCACTGCCTGACGGGCCTTGACCTGGCCGCCC  1133

 146  ...........GlyPro................ProPro  1149
      |||||                |||||
1134  TCAAGG.......................CCGCAGCCCGAGGGCTCTTC  1159
      :::|||                   |||||
1150  GlyArgTyrTrpSerValAspLysTrpLeuGlyGlyTrpArgAlaGlySerIr  1166

1160  CTGCTACTGCGCCACTGGGACCAAAACCTGCAGCTACACCTGC......T  1203
      ::: :::::::|||          ||| |
1166  pAspTyrLeuProProArgSerGlyProAlaAlaTrpHisCysArgHisC  1183

1204  GTGCTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCTGCTGTGGGCTG  1253
      ||||::::  |||            ||| |||
1183  ysAlaSer........................LeuGluLeu  1188

1254  GCGCCCACCCAACACACTCAAGTCACTGCGCGCCCAGGGCTGGCCTCTTG  1303
      ||| ||| ||| :::  |||||
1189  LeuProProArgHisLeuSerCysSerHisAspGlyLeuAspGly..  1204

1304  GTGCTGGGAAGTGTAGGCTGGTGCGCCAGCCTGTCCCCCACTGCTTCTTAC  1353
      |||||| |||  |||
1205  ...........GlyTrpTrpAlaProProPro............  1211

1354  TCCCTCCCTGGAGC............CCTCTTGCCCCCCACAAAAAGTGCCTGCCGCCTG  1397
      ||||||||:::          :::::  |||
1212  ..ProProTrpArgAlaAlaGlyProLeuProArgArgArgAlaArgCysGly  1227

1398  TGC......TCTCTCCCTCCTCCCCACCCCATCACACTGCTTCAC  1441
      ||| |||  ||| |||:::
1228  CysProArgSerHisProHisArgProHisArgProAlaSerHisArgThrProAl  1244

1442  CTCTGAGCTCCCT  1454
      ::: |||
 244  aAlaAlaAlaPro  1248

seq_documentation_block:
ID   TEGU_EBV     STANDARD;      PRT;    3149 AA.
AC   P03186;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Large tegument protein.
GN   BPLF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
```

seq_name: SwissProt_40:TEGU_EBV

RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR   PIR; A03742; QQBE3.
KW   Hypothetical protein; Early protein; Repeat.
FT   DOMAIN   149   648   4 X 125 AA TANDEM REPEATS.
FT   REPEAT   149   273   1.
FT   REPEAT   274   398   2.
FT   REPEAT   399   523   3.
FT   REPEAT   524   648   4.
SQ   SEQUENCE   660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

alignment_scores:
            Quality: 196.50   Length: 480
            Ratio: 0.949   Gaps: 30
Percent Similarity: 43.125   Percent Identity: 27.708

alignment_block:
US-09-786-136-4 x YHL1_EBV   ..

Align seg 1/1 to: YHL1_EBV from: 1 to: 660

350 CCTGGGGCCCTGCAGGTGTGCAGC........................TCAGAGCCGGCTGA 387
     |||||||| ::: ||| :::::||||||
172 ProGlyAlaGlyGlnArgProSerGlyProThrGlyArgProAlaAl 188

388 GCCC.....CCAGGAACCCTCCCTTGCCTCTCACAGCCATCTAGATGCAG 431
     |||| |||||||||||| |||
188 aProGlyAlaProGlyThrPro........................AlaA 197

432 CTCCTCTGCCCACTGTTGTCTACTCTAAAGGACTTCAGAGAGGCTCTCCA 481
     |||||||| || |||
197 laProGlyPro..................GlyGlyGlyAlaAlaValPro 207

482 GCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCGTGCTTTGGG 531
     ::::|||||| :::||::::::|
208 SerGlyAlaThrProHisProGluArgGlySerGly.............. 219

532 GGACCCTGCCACTCCACGGAAGGTCCTCCGCGCCCACCT........... 571
     |||||| ||| :: ||| ||||||
220 ....ProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnG 235

572 ..CCCCGTCCTCCCTGTCGGCTCAGCATGGCCGCCGTCACAAACTCTGT 619
     ::::::|| |||||||:||
235 luProArgLeuProGlnAspProLeuAlaAlaGlnArg.........Cys 248

620 .....AGCCCTGACCCGGCCCAGGCCAACAACAGTGAGGCAGGCAGCCACCA 663
     ::|||| ||| :::|||| ||:||||
249 ProAlaGlyProProPro.ThrArgSerGlyAlaAlaAlaGlnArgThrH 265

664 TGACTACCTGCCCTGGTGC........................GGCTGC... 689
     ||| ||||||
265 isArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysPro 281

690 ................................AGGAGGCACCAGGCTCCTTC 709
     ::: ||| ||| ::||||||
282 ArgThrTrpArgArgSerGlyAlaGlyHisProProProGl 298

---

410 : ||| :|||||||| ::: |
     nThrProLysArgLysLysGlyLeuGlyLysAspSerProHisGlyLysP 427

979 CGACCTGCGGGGCCGGACCTGGGCCACTGCCGTAGCACTGCCGCCGGCTGG 1028
     |||| ||| |||
427 roThrSerGlyArg........ArgLeuProLeuSerThrArgThrAsp 440

1029 AGCACCGATGCGCCGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAG 1078
     ::: ||| ||||||| ||| ::|||||
441 ThrGluAspGlnLeuProArgThrHisValProProHisArgProPr 457

1079 GCTGATTGCTGGCTGCGGGGCCCCAGCACTTGCCTGACGGCCTTGACCTGGC 1128
     ||| ||| ::: ||| ||||
457 oSerAlaAlaArgLeuProProValIleProIleProHisGlnSerP 474

1129 CGGCCCTCAAGGCCGCAGCCCGAGG..................... 1152
     |||::: ||| ||| ::|||:::
474 roProAlaSerPro.ThrProHisProAlaProValSerThrIleAlaPr 490

1153 ....GCTCTTC 1159
                                              |||
490 oSerValThrProSerProArgProLeuGlnIleProIleProLeuP 507

1160 CTGCTACTGCGCCACTGGACCAAAA...CCTGCAGCTACACCTGCTGTG 1206
     || :::|||||| ::|||||| || ::||||||:::
507 roGlnAlaAlaProSerAsnProLysIleProLeuThrThrProSer... 522

1207 CTACAGCCCAGCGAACGTGTGAAGGCTGCCCCCTGCTGTCGTTGGGCTGGCG 1256
     ||||| ::: |||||| ||: :: |||
523 .....ProSerProThrAlaAlaAlaAlaAla.ProThrThrThrThrLeuS 537

1257 CCCCACCCACACACTCAAGTCACTGCCGCCCAGGGCTGGCCTCTTGGTG 1306
     ::|||||| ||| ::::: ||||||||||:::::::
537 erProProProThr...GlnGlnGlnProGlnProGlnSerAla........ 549

1307 CTGGGAAAGTGTAGGCTGCTGGTGCCCAGCCTGTCCCCACTGCTT....... 1348
     :::|||||| ||||||||
550 ....AlaProAlaProSerProLeuLeuProGlnGl 560

1349 ...CTTACTCCCTCCTGGAGCCCTCT...................... 1372
     |||||||| |||:::
1349 560 nGlnProThrProSerAlaAlaProAlaProSerProLeuLeuProGlnG 577

1373 ....TGCCCCCAC 1381
                                     |||:::::
577 lnGlnProProSerAlaAlaArgAlaProSerProLeuProProGln 593

1382 AAAAAGTGCCTGCCTGCTGCTCCTCCTCCTCCTCCACCCCCACTCCACACT 1431
     ::::: |||||| :::||||| ||||| :::::
594 GlnGlnProLeuProSerAlaThrProAlaPro.ProProAlaGlnGlnL 610

1432 CCCCTCCATCCTCTGAGCTC....................CCTGCA 1457
     |||||||||::::: ||||||
610 euProProSerAlaThrProThrLeuGluProGluLysAsnHisProProAla 626

1458 ACACAGTGGGAAGGGTAGAGAGCCACAGTCCCAAATCCT 1496
     ::::: ||| ||| ||||||:::|||
627 AlaAspProArgAlaGlyThrGlu...IleSerProSerPro 638

seq_name: SwissProt_40:YHL1_EBV

seq_documentation_block:
ID   YHL1_EBV   STANDARD;   PRT;   660 AA.
AC   P03181;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical BHLF1 protein.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]

```
                      |||  :::           |||        :::::::::|||
589  ProGluArgGlySerGlyPro.AlaAspProProAlaAlaAlaArgLeuP  605

1453 CTGCAACACAGTGGAAGGGTAGAGAGCCACAG 1484
      :::   |||   ||||||
605  roProGluArgGlnGluProArgLeuProGln 615
```

```
710  CGCCTGGACGGCGCCCTTCGCCGCGTGTGCCGCATCTCGCAGGAGCGCC.  758
     :|||   ||||||||:|||    ::::||||  :::|||
298  yAlaGlyGlnArgProSerGlyPro...ThrGlyGlyArgProAlaAlaP  314

758                                                      758
314  roGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla     330

759  .......TCTGCCGTGCCTCGCCCTTG.........CCGTGCAC        787
       ::|||:||||:|||:|||         |||:||
331  ValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAs  347

788  CGCGCCAGCCTCAGCC......CCAACCTCGGCCTCATTGCCCTGGCACT  831
     ||||||||  ||||||||:|||         :::  |||
347  pProProAlaAlaArgLeuProProGluArgGlnGluProProArgLeuP  364

832  TCTGGGCCCTGGTGTTGGCCAGGTGACAGTGCCACGGCCCTCCTGCCAGCC 881
     ||||| |||:|| |||::|||      ||||||     :::
364  roGlnAspLeuAlaAlaGlnAlaGlnArgCysProAlaGlyProProThr  380

882  CGTCCCCCAGCTCGGGCTCGAGGGGCCAGGCCAGG...........       917
     |||  ||| ||| ||| ::: |||
381  ArgSerGlyAlaAlaAlaGlnArgThrHisArgProProGlyCysPr     397

918  ...................................TGG...ACAGTGGGCG  930
                                        |||     :::|
397  oArgSerAlaArgAsnProGlyCysProArgThrTrpArgArgSerG    414

931  GGGCTCAGCACACCGAGGCCTCCGATGGGGCGG...AAGGGCTGGGCGGCA  977
     ||||||| ::|||||  |||    ||||||    :::::::
414  lyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGly  430

978  CCGACCTCGGG.....GCCGACCTGGG...CCACTGCCGTAGCACTC     1018
     |||||| |||    ||::||||||||| |||          ::
431  ProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAl  447

1019 GCCTGGCTGGAGCACCGATGGCCGCGCTGCCTTCGACGAGTGGGAACTGAC 1068
     :||||||    :::  |||  ::|||||||
447  aProGlyProGlyGlyGlyAlaAlaValProSerGlyAla.........   460

1069 AGCGGCCAAGGCTGATTGCTGGCTGCGCGGGCCCAGCACTTGCCTGACGGCC 1118
     |||::::          ||||||:||:|||      ::|||
461  ..ThrProHisProGluArgGlySerGlyProAlaAspProProAlaAla  476

1119 TTGACCTGGCCGCCCTCAAGG..............CCGCAGCCCGAGGG    1153
     ||||||:::|||              ||||||     ::
477  AlaArgLeuProProGluArgGlnGluProGlnAspLeuAl          493

1154 CTCTTCCTGCTACTGCCGCCATGGACCAAAACCTGCAGCTACACCTGCT   1203
     ::::::  ||| ||| |||||| ::: |||   |||
493  aAlaAlaGlnArgCys.ProAlaGlyProProThrArgSerGlyAla    509

204  GTGCTACAGCCCAGCGAACGT.............GTGAA           1229
     ::: |||  ::::||||            |||||   :::::
510  AlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaAr  526

1230 GGCTGCCCCCTGCGCTGGCTGGCTGGCGCCCCAACACACTCAAGTCA     1279
     ;    ||||||        :::  :::|||
526  gAsnProGlyCysProArgThrTrpArgArgSerGlyAlaGln....   541

1280 CTGCCGCCCAGGGCTGG...CCTCTTGGCTGCTGGGAAAGTGAGGCTGGT 1326
     :::|||   |||  ||| ||||||||:::       :::|||
542  ......ArgGlyHisProProGlyAlaGlyGlnArgProSerGly     555

1327 GCCAGCCTGTCCCCCACGTCTTCTTACTCCCCTGGAGCCCTCTTGCC     1376
     ::: :::  ||| |||||||||:::      ||||||||
556  ProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAl 572

1377 CCCA.............CAAAAAGTGCCTGCCTGTGCTCTCCCCTCT     1411
     |||             ||||||:::|| |||||||::|||
572  aProGlyProGlyGlyGlyAlaAlaValProSerGlyAla.ThrProHis 588

1412 CCTCCCACC.........CCACTCACACTCCCCTCCATCCTCTGAGCTCC  1452
```

OM of: US-09-786-136-4 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 19, 2002 5:42 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US09786136/runat_16092002_141242_19186/app_query.fasta_1.1592
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-ELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-NS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-MAX=100 -THR_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs
-M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09786136_@CGN1_1_14 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-136-4
Query length: 1511
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 2425594
Search time (sec): 34.210000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-026-138E-4	+	205.50	248.07	1.8e-07	1323	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-483-533-41	+	193.50	245.10	1.0e-06	355	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-283-471A-41	+	193.50	245.10	1.0e-06	355	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US91-06532-3	+	193.50	245.10	1.0e-06	355	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-231-193A-58	-	191.00	228.93	2.1e-06	1336	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-486-273A-58	-	191.00	228.93	2.1e-06	1336	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-940-086A-58	-	191.00	228.93	2.1e-06	1336	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-940-035A-58	-	191.00	228.93	2.1e-06	1336	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-316-650-12	-	189.50	226.22	2.8e-06	1442	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-02251-12	-	189.50	226.22	2.8e-06	1442	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-906-865-4	-	189.00	232.64	2.5e-06	696	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-931-820-3	-	188.50	227.90	3.1e-06	1060	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-20	-	188.50	225.07	3.3e-06	1418	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-010-999-1	-	188.50	225.07	3.3e-06	1418	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-570-573-20	-	188.50	225.07	3.3e-06	1418	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-20	-	188.50	225.07	3.3e-06	1418	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-945-283-2	-	187.00	219.96	4.6e-06	1958	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-313-288B-15	-	182.00	227.29	7.5e-06	469	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-209-747-2	-	179.50	218.43	1.3e-05	832	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-458-298-2	-	179.50	218.43	1.3e-05	832	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-103-429A-4	+	178.50	217.44	1.5e-05	805	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-556-978B-19	-	177.00	217.53	1.9e-05	651	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-247-806-1	-	177.00	217.53	1.9e-05	651	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-09-034-177-3	-	177.00	216.58	1.9e-05	718	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-425-069-2	-	177.00	216.19	2.0e-05	747	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-49	-	175.50	213.29	2.6e-05	822	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-628-829-4	-	174.50	205.54	3.6e-05	1593	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-660-963-12	+	174.00	216.19	2.9e-05	498	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-5	+	173.50	209.15	3.7e-05	960	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-556-978B-63	-	173.00	213.00	3.6e-05	604	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-556-978B-21	-	173.00	211.66	4.3e-05	606	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-247-806-4	-	172.00	211.66	4.3e-05	606	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-4	+	172.00	209.98	4.5e-05	720	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-642-255-53	+	172.00	209.24	4.6e-05	777	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-046-585-5	+	172.00	199.88	5.8e-05	2035	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-393-703-5	+	172.00	199.88	5.8e-05	2035	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US93-11721-5	+	172.00	199.88	5.8e-05	2035	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-556-978B-59	-	171.00	212.35	4.8e-05	493	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-19	-	171.00	202.44	6.2e-05	1366	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-19	-	171.00	202.44	6.2e-05	1366	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-19	-	171.00	202.44	6.2e-05	1366	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-247-806-6	-	170.50	209.69	5.5e-05	606	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79	-	168.00	199.65	0.0001	1213	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-931-820-2	-	167.00	199.99	0.0001	1024	

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-026-138E-4

seq_documentation_block:
; Sequence 4, Application US/08026138E
; Patent No. 5502166
;   GENERAL INFORMATION:
;   APPLICANT: Masayoshi MISHINA
;   TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nishiohata Residence 1-107
;   STREET: 5214, Nishiohata-machi
;   CITY: Niigata-shi
;   STATE: Niigata-ken
;   COUNTRY: JAPAN
;   ZIP: 951
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS v.5
;   SOFTWARE: Word Perfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/026,138E
;   FILING DATE: 26-FEB-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 39563/1992
;   FILING DATE: 26-FEB-1992
;   APPLICATION NUMBER: JP 173155/1992
;   FILING DATE: 30-JUN-1992
;   APPLICATION NUMBER: JP 215017/1992
;   FILING DATE: 12-AUG-1992
;   APPLICATION NUMBER: JP 303878/1992
;   FILING DATE: 13-NOV-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hamburg, C. Bruce
;   REGISTRATION NUMBER: 22,389
;   REFERENCE/DOCKET NUMBER: F-4551
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 986-2340
;   TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1323 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single strand
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: mouse
;   TISSUE TYPE: brain
;   PUBLICATION INFORMATION:
;   AUTHORS: Masayoshi MISHINA
;   TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
;   RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 to 1323
US-08-026-138E-4

alignment_scores:
Quality: 205.50   Length: 443
Ratio: 1.309   Gaps: 26
Percent Similarity: 35.440   Percent Identity: 26.185

alignment_block:
US-09-786-136-4 x US-08-026-138E-4 ..
Align seg 1/1 to: US-08-026-138E-4 from: 1 to: 1323

353 GGGGCCCTCGAGGTGTGCAGCTCAGGCCCGCTGAGCCCCAGGAACCCC 402
    ||| ::: |||||:::||| ||| |||||:::||| |
889 GlyMetTyrSerCysSerAlaGluAlaAlaProToProAlaLysPr 905

403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCCTGCCCACTGTTGTCT 452
    |||| |||||||||:::|||
905 oProPro..........ProProGlnProLeuProSerProAlaT 917

453 ACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGCGCCTGGGACTCGGAC 502
    ||||||
917 yr..........ProAla.......... 919

503 CAAAATGGCAACTCCAAGCGTGCTTTGGGGACCCTGCCACTCCCACGA 552
    ||| |||
920 ....AlaArgProProPr 924

553 AGGTCCTCGCCGCCCACCTCCCCGTCCTCCCTGCTCGGCTCAGCATGGGC 602
    |||||| || ||||||
924 oGlyProAlaProPheValProArgGluArgAlaAlaAlaAspArgTrpA 941

603 GCCGTCACAAACTCGTAGCCTGACCCGGGCCCAGGCCAACAACAGTGAA 652
    |||| ||| :::||| |||||| ||| :::|||
941 rgArgAlaLysGlyThrGlyPro..ProGlyGlyAlaAlaLeuAlaAsp 956

653 GGCAGCGACCATGACTACCTGCCCTTGGTGCGGCTGCAGGAGGCACCAGG 702
    ||||||||| ::: ||| ||| ||| ||| :::
957 GlyPheHisArgTyrTyrGlyProIle..........GluProGlnGl 969

703 CTCCTTCCGCTGGACGCGCCCTTCTGCCCGCTGTGCCGCATCTCCGCAGG 752
    -   | |||||| :::|||||||
969 yLeuGlyLeuGlyGluAla.ArgAlaAlaPro..........Arg 980

753 AGGGCCTCTGCGTGCCTCCCCCTTTGCCGTGCACCGCGCCAGCCTCAGC 802
    :::|||:::||: ||| ||| ||| :::
981 GlyAlaAlaGlyArgProLeuSerProProThrThrGlnProProGlnLy 997

803 CCCACCTCGGCCTCATTGCCCGGCACTTCTGGGCCTCTGGTGTTGGCCA 852
    |||||||||:::||| |||:::||| ||||||
997 sProPro..........ProSerTyrPhe..AlaIleValArgGluG 1009

853 GG.......... GTGACAGTG 863
    ::
1009 lnGluProAlaGluProProAlaGlyAlaPheProGlyPheProSerPro 1025

864 CCACGGCCTCCTCCAGCC.......... 881
    ||| ||||||||||
1026 ProAlaProProAlaAlaAlaAlaAlaAlaAlaAlaAlaValGlyProProLeuCysAr 1042

882 ....CGTCCCCAGCT.......... 900
    ||||||||| ||| ||| :::
1042 gLeuAlaPheGluAspGluSerProSerProAlaSerAlaGlyArgValL 1059

01 TGAGGGGCCAGGCC...AGGTGGACAGTGGGCGGG..........GCT 935
   || :::||| ||| |||||:::
59 euThrProArgAlaSerArgCysTrpValGlyAlaArgAlaAlaArgAla 1075

936 CAGACACCGAGGCCTCCGATGGGCGGAAGGCTGGGCGGCACCGACCTG 985
    ||||||||| ||| ||| || ||
1076 LeuGlyProArgProHisHisArgArgValArgThrAlaProProProCy 1092

986 C.......... 986
    ||
1092 sAlaTyrLeuAspLeuProSerProSerAspGluSerGlus 1109

987 ...GGGGGCCAGGACCTGGGCCCACTGCCGTAGCACTCGCCGGCTGG 1028
    ||||||::: ||| ||| :::|||||||
1109 erLeuGlyAlaSerLeuGlyGlyLeu.......GluProTrpGTrp 1122

1029 AGCACCGATGCGCCCCGTCCGCTTCGACGAGTGGGAACTGACAGCGGCCAAG 1078
    :::||| ||| |||:::|||
1123 PheAlaAspPhePrToTyrProTyrAlaGlu..........Ar 1133

1079 GCTGATTGGCTGGCTGCGCGGGCCCAGCACTTGCCTGACGACGGCCTTGACCTGGC 1128

1133 ||||| ||||||
    gLeu..........GlyPro 1137

1129 CGCCTCAAGG 1139
    |||||:::|||
1137 roProGlyArgTyrTrpSerValAspLysLysGlyGlyTrpArgAlaGly 1153

1140 .... CCGGCACCCGAGGGCCTCTTCCTGCTGCTACTGCGCCAC 1174
1154 SerTrpAspTyrLeuProProArgGlyGlyProAlaTrpHisCys.... 1168

1175 TGGGACCAAAACCTGCAGCTACACCTGCTGTGCTACAGGCCAGCGAACGT 1224
    ||| |||
1169 ....ArgHisC 1171

1225 GTGAAGGCTGCCCCTGCTGCTTGGGCTGGCGCCGCCCACCCAACACACTCA 1274
    |||::: |||||| ||| |||
1171 ysAlaSerLeuGluLeuLeu..........ProProProArgHisLeu 1183

1275 AGTCACTGCCGCCAGGCTGGCCTCTTGGTGTGCGTGGGAAAGTGTAGGCTG 1324
    ||| ::: ||||| ||||||
1184 SerCysSerHisAspGlyLyLeuAspGly..........GlyTr 1194

1325 GTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCCCTGGAGC......C 1368
    |||||| ||| || ||||||:::
1194 pTrpAlaProProPro..........ProProTrpProAlaAlaGlyP 1206

1369 CTCTTGCCCCACAAAAATGCCTGCCGTGCTCTCCCCTCCTCCCTCCTCCCA 1418
    | |||||| :::||| ||
1206 roProAlaProArgArgAlaArgCysAlaArgCysGlyCysPro......... Arg 1218

1419 CCCCACTCACACTCCCCTCCATCCTCT 1445
    ||||| |||||| |||:::
1219 ProHisProHisArgProArgAlaSer 1227

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-483-533-41

seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No. 617047
; GENERAL INFORMATION:
;   APPLICANT: Roizman, Bernard
;   APPLICANT: Chou, Joany
;   TITLE OF INVENTION: Method for Treating Tumorigenic
;   TITLE OF INVENTION: Diseases
;   NUMBER OF SEQUENCES: 43
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/483,533
;   FILING DATE: 07-MAR-95
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/419,853
;   FILING DATE: 11-APR-95
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/861,233
;   FILING DATE: 31-MAR-92
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Zeller, James P.
;   REGISTRATION NUMBER: 28,491
;   REFERENCE/DOCKET NUMBER: 28097/32742
;   TELECOMMUNICATION INFORMATION:

```
;   TELEPHONE: 312/474-6300
;   TELEFAX:  312/474-0448
;   TELEX:    25-3856
; INFORMATION FOR SEQ ID NO: 41:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:   355 amino acids
;   TYPE:     amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-483-533-41

alignment_scores:
        Quality:    193.50          Length:      378
        Ratio:      1.166           Gaps:         22
cent Similarity:  43.915    Percent Identity:  28.571

alignment_block:
US-09-786-136-4 x US-08-483-533-41

Align seg 1/1    to: US-08-483-533-41  from:  1  to:  355

353 GGGGCCCTGCAGGTGTGCAGTCTCAGAGCCGCCTGAGCCCCCAGGAACCCC 402
    ||||||:: :::::::||||::::::|::|:: ||:::|||::::||:::::
 20 GlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSerGl 36

403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCTGCCC..........  442
    ||||||| ||:||| ||||||||||||||||||:: :||
 36 uProAlaValArgSerAlaProAlaAlaAlaProProProProProAlaG 53

443 ...........ACTGTTGTCTACTCTAAAGGACTTCAGAGA 472
               :::::::::::    ::: |||:|||
 53 lyGlyProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis.... 68

473 GGCTCTCCAGCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCG 522
    ||| ||:||| ||:|||:||:||:|::|||:||:::     ::
 69 ...ValProGluSerAlaSerAspAspAspAspAspAspTrpProAs 84

523 TGCTTGGGGGACCCTGCCACTCCCACGAAGTCCTCGCCGCCCACCTC 572
    ::: ||||||| ||:::::||   |||:|||||
 84 pSerProProGluSerAlaProGluAlaArgProThrAlaAlaAlaP 101

573 CCCGTCCTCCCCTGCGGCTCAGCATGGGCCGCCGTCACAAACTCTGTAGC 622
    |||||||
101 roArgPro.......................................... 103

623 CCTGACCCGGGCCAGGCCAACAACAGTGAAGGCAGCGACCATGACT.... 668
    ||| |||  |||||||:::  ::::::|||  ||:|||
104 ProGlyPro.HisArgProAlaTrpAlaArgGlyAlaGlyLeuThrProP 120

669 ..ACCTGGCCCTTGGTGCGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGG 716
    ::::::::::::|||||::::
120 roThrProPro...........ArgAlaProSerAlaPhe 129

717 ACGGGCCCCTTCTCGGCCGCGTGTGCGCATCTCGCCAGGAGGCGCCTGCCGT 766
    |||  ||||||||||| |||:|||||||||:::
130 ArgArgAlaSerProSerAlaCysAlaSerProArgSerArgThrTrp... 144

767 GCCTCGGCCCTTTGCGTGCACCGCGCAGCCTCA..........GCCC 804
    |||  ||| |||| |||::        :::|
145 ...ArgAlaCysAspAlaCysAlaArgAlaGlyArgArgGlyArgSerP 160

805 CACCTCGGCCTCATTGCCCTGGGCACTTCTGGGCCCTGGTGTTGGCCAGG 854
    ||||||||||   |||
160 roProArgProProArgPro.............................. 166

855 GTGACAGTGCCACGGCCTCCT...GCAGCCCGTCCCCCA.......... 890
    |||||||||||          ||||||||
167 .......ProArgProProArgProProProArgGlyCysAl 180

891 .GCTCGGGCTCTGAGGGGCCAGGCCAGTGGACAGTGGGCGGGGCTCAGA 939
    :::|||       ||| |||   |||                   |||
```

```
180 aSerArgProThrSerGlyCysAlaThrTrpSerGly........... 193

940 CACCGAGGCCTCCGATGGGCCGGAAGGGCTGGGCGGCCACCGACCTGCGGG 989
    |||||||||||   ||| |||||||||        :::||
194 ..ProArgProProAlaTrpArg.AlaAlaAlaArgGlyProAlaSerGl 209

990 GCCG...GACCTGGGCCACTGCGCTAGCACTCGCCTGCTGGAGCACCGA 1036
    ||||   |||||||       |:||||||
209 yProThrGlyLeuGly..............SerGlyAlaGlyT 219

1037 TGCGCCGCTGCCTTGACGAGTGGGAACTGACGCGGCGCCAAGGCTGATTG 1086
     |||||   ||| |||    |||||
219 rpArgArgProArgArgSerGlyArg....................... 228

1087 CTGGCTGCGGGCCCAGCACTTGCCTGACGGCCCTTGACCTGGCCGCCCTCA 1136
     |||||  ||||||    ::  |||||    |||  |::::|::::|||
229 ...AlaTrpGlyProArgProValProGlyProTrpProAlaGluProAl 244

1137 AGGCCGC...........................AGCCCGAGGGCT.....CTTCCTGCTA 1165
     ||:: ||| :::                AGCCCGAGGGCT      ::
244 aArgArgThrArgSerAsnValThrProGluAlaAlaTrpValPheArgG 261

1166 CTGGCTGCGGGGACCAAAAAACCTGCAGCTACACCTGCTGTGCTACAGCCC 1215
     |||||  |||||     ::  |||||||     ::|||
261 lyAlaProGlySerSerAlaProSerArgSerProGluArgArgTrpGln 277

1216 AGGCAACGTGTGAAGGCTGCCCCCTGCGTCGTCGGCGTGGCGCCCCACCCA 1265
     ::: ||||::   ||||||:  |||||||       ::
278 GluProArgIleTyrThr...............LeuGlyAlaSerProProSe 290

1266 ACACACTCAAGTCACTGCCGCCCAGGGCTGGCCTCTTGGTGCTGGGAAAG 1315
     ||||||       :::::: |||||||||::|||||||
290 rGlnGlyGlyProProArgGlyArgAspTrpProIleGly...GlyArgG 306

1316 TGTAGGCTGGTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCCCTGGA 1365
     :::|||||   ||||||||||||:::  |||||||
306 lnArgGlyGlyAla.......ArgProThrSerValArgArgValPheGly 320

1366 GCCCTC...............TTGCCCCCA 1380
     |||    |||||||||
321 AlaAlaProIleGlyArgGlyArgGluLeuProPro 330

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-283-471A-41

seq_documentation_block:
; Sequence 41, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
;   APPLICANT: Roizman, Bernard
;   APPLICANT: Chou, Joany
;   TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
;   NUMBER OF SEQUENCES: 43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;     STREET:    6300 Sears Tower, 233 South Wacker Drive
;     CITY:      Chicago
;     STATE:     Illinois
;     COUNTRY:   United States of America
;     ZIP:       60606-6402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER:    IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE:    PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/283,471A
;     FILING DATE:  04-APR-1999
;     CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/861,233
;     FILING DATE:  31-MAR-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/419,853
```

```
;     FILING DATE:   11-APR-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:   08/483,533
;     FILING DATE:   07-JUN-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:   Zeller, James P.
;     REGISTRATION NUMBER:   28,491
;   TELECOMMUNICATION INFORMATION:
;     REFERENCE/DOCKET NUMBER:   27373/32742A
;     TELEPHONE:   312/474-6300
;     TELEFAX:   312/474-0448
;     TELEX:   25-3856
; INFORMATION FOR SEQ ID NO:  41:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:   355 amino acids
;     TYPE:   amino acid
;     STRANDEDNESS:   single
;     TOPOLOGY:   linear
;   MOLECULE TYPE:   protein
US-09-283-471A-41
```

```
ment_scores:
      Quality:   193.50          Length:    378
      Ratio:     1.166           Gaps:      22
Percent Similarity:   43.915    Percent Identity:   28.571

alignment_block:
US-09-786-136-4 x US-09-283-471A-41

Align seg 1/1  to: US-09-283-471A-41  from: 1  to: 355

353 GGGGCCCTGCAGGTGTGCAGTCAGAGCCCCTGAGCCCCCAGGAACCCC 402
20  GlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSerGl 36

403 TCCTGCCTCTCACAGCCATCTAGATGCAGTCCTCTGCCC.......... 442
36  uProAlaValArgSerAlaProAlaAlaAlaProProProProAlaG 53

443 ............ACTGTTGCTACTCTAAAGGACTTCAGAGA 472
53  lyGlyProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis... 68

473 GGCTCTCCAGCAGGCGCCTGGGACTCGGACCAAAATGGCAACTCCAAGCG 522
68  ..ValProGluSerAlaSerAspAspAspAspAspAspTrpProAs 84

523 TGCTTTGGGGGACCCTGCCACTCCCACGGAAGGTCCTCCGCCGCCCACCTC 572
84  pSerProProGluSerAlaProGluAlaProThrAlaAlaAlaAlaP 101

573 CCCGTCTCCCCTGTCGGCTCAGCATGGCCCGCCGTCACAAACTCTGTAGC 622
101 roArgPro.......... 103

623 CCTGACCGGGCAGGCCAACACAGTGAAGGCAGGACCATGACT.... 668
104 ProGlyPro.HisArgProAlaTrpAlaArgGlyAlaGlyLeuThrProP 120

669 ..ACCTGCCCTTGGTGCGGCTGCAGGAGGCACCCAGGCTCCTTCCGCCTGG 716
120 roThrProPro............ArgAlaProSerAlaPhe 129

717 ACGCGCCCTTCTCGCCGGTGCCATCCGCAGGAGCCGCCTCTGCCGT 766
130 ArgArgAlaSerProSerAlaCysAlaSerProArgSerThrTrp...... 144

767 GCCTCGCCCTTTGCCGTGCACCGCGCCAGCCTCA..........GCCC 804
145 ...ArgAlaCysAlaCysAspAlaArgAlaGlyAlaGlyArgGlyArgSerP 160

805 CACCTCGGCCCTCATTCCCCTGGCACTTCTGGCCCCTGGTTGGCCAGG 854
```

```
160 roProArgProProArgPro.......... 166

855 GTGACAGTGCCACGGCCTCCT..GCAGCCGTCCCCA.......... 890
167 ..........ProArgProProArgProProArgProProArgGlyCysAl 180

891 .GCTCCGGCTCTGAGGGCCAGGCCAGGTGGACAGTGGGCGGGGCTCAGA 939
180 aSerArgProThrSerGlyCysAlaThrTrpTrpSerGly.......... 193

940 CACCGAGGCCTCCGATGGGGCGGAAGGCGTGGGCGCCGCACCGACCTGCGGG 989
194 ..ProArgProProAlaTrpArg.AlaAlaAlaArgGlyProAlaSerGl 209

990 GCCG...GACCTGGCCACTGCGTAGCACTCGCCTGGCTGGAGCACCGA 1036
209 yProThrGlyLeuGly..........SerGlyAlaGlyT 219

1037 TGCGCCGCTGCCTTGCACGAGTGGAACTGACAGCGGCCAAGGCTGATTG 1086
219 rpArgArgProArgSerGlyArg.......... 228

1087 CTGGCTGCGGGCCCAGCACTTGCCTGACCGGCCTTGACCTGCCGCCCTCA 1136
229 ...AlaTrpGlyProValProGlyProrpProAlaGluProAl 244

1137 AGGCCGC..........AGCCCGAGGGCT.....CTTCCTGCTA 1165
244 aArgArgThrArgSerAsnValThrProGluAlaAlaTrpValPheArg 261

1166 CTGGCCACTGGGACCAAAACCTGCAGCTACACCTGCTGTCACAGCCCC 1215
261 lyAlaProGlySerAlaProSerAlaProSerArgSerProGluArgArgTrpGln 277

1216 AGCGGAACGTGTGAAGGCTGCCCCCTGCTGCTTGGGCTGGCGCCCACCCA 1265
278 GluProArgIleTyrThr..........LeuGlyAlaSerProProSe 290

1266 ACACACTCAAGTCACTGCCCGCCCAGGGCTGGCCTTGGTGCTGGGAAAG 1315
290 rGlnGlyGlyProProArgGlyArgAspTrpProIleGly...GlyArgG 306

1316 TGTAGGCTGGTGCCAGCCTGCTCCCCACTGCTTCTTACTCCCTCCCTGGA 1365
306 lnArgGlyGlyAla..........ArgProThrSerValArgArgValPheGly 320

1366 GCCCTC..........TTGCCCCA 1380
321 AlaArgProIleGlyArgGluLeuProPro 330
```

```
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/06532
;       FILING DATE:  19910910
;       CLASSIFICATION:  424
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Gruber, Lewis S.
;       REGISTRATION NUMBER:  30,060
;       REFERENCE/DOCKET NUMBER:  27373/8235
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/346-5750
;       TELEFAX:  312/984-9740
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  3:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH:  355 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
PCT-US91-06532-3

alignment_scores:
        Quality:       193.50          Length:      378
        Ratio:         1.166           Gaps:         22
Percent Similarity:    43.915    Percent Identity:  28.571

alignment_block:
US-09-786-136-4  to:  PCT-US91-06532-3   .

Align seg 1/1  to:  PCT-US91-06532-3   from: 1  to: 355

353 GGGGCCCCTGCAGGTGTGCAGCGTCAGAGCCCCGCTGAGCCCCCAGGAACCCC    402
    ||||||::  :::::::::||::: ::::::: ||::::||
 20 GlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSerGl      36

403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCCTCGCCC...........    442
    ||||||  :::|||     |||||||||  |||
 36 uProAlaValArgSerAlaProAlaAlaAlaProProProProProProAlaG  53

443 ...........ACTGTTGTCTACTCTAAAGGACTTCAGAGA             472
               :::::::|||     :: |||:::
 53 lyGlyProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis...       68

473 GGCTCTCCAGCAGGCGCTGGGACTCGGACCAAAATGGCAACTCCAAGCG      522
    |||  ::|||  |||::|||:::||::::: :::
 69 ...ValProGluSerAlaSerAspAspAspAspAspAspTrpProAs       84

523 TGCTTTGGGGACCCTGCCACTCCCACGGAAGGTCCTCGCCGCCCACCTC     572
    ::::        :::::|||    ||  |||:::
 84 pSerProProGluSerAlaProGluAlaArgProThrAlaAlaAlaP      101

573 CCCGTCCTCCCCTGTCCGGCTCAGCATGGCCGCCGTCACAAACTCTGTAGC    622
    ||||||||
101 roArgPro.......................................    103

623 CCTGACCCGGCCAGGCCAACACAGTGAAGGCAGCCACCATGACT.....    668
    |||  |||    |||||::   :::::::::|| ::||
104 ProGlyPro.HisArgProAlaTrpAlaArgGlyAlaGlyLeuThrProP   120

669 ..ACCTGCCCTGTGCGCGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGG    716
    |||  |||             ::||||||||||||:::
120 roThrProPro...............ArgAlaProSerAlaPhe       129

717 ACGGCGCCCTTCTCGCCCGCTGTGCCCATCTCGCCAGGAGGCGCCTCTGCCGT   766
    |||     |||         ||||||||::
130 ArgArgAlaSerProSerAlaCysAlaSerProArgSerThrArg.....    144

767 GCCTCGCCCCTTTGCCGTTGCACCCGCCGCAGCCTCA...........GCCC    804
    ||| ||| |||   ||:::          :::||
145 ...ArgAlaCysAlaCysAspAlaArgAlaGlyLyrArgGlyArgSerP     160

805 CACCTCGGCCTCATTGCCCTGGGCACTTCTGGGCACTTCTGGTGTTGGCCAGG    854
    ||||||||||         |||
```

```
160 roProArgProProArgPro.........................    166

855 GTGACAGTGCCACGGCCTCCT..GCAGCCCGTCCCCA             890
    |||||||||||                ||||||||
167 .........ProArgProProArgProProArgGlyCysAl          180

891 .GCTCGGCTCTGAGGGCCCAGGTGACAGTGGCGGGGGCTCAGA          939
    ::||| ||| |||  |||::::::||:|||
180 aSerArgProThrSerGlyCysAlaThrTrpSerGly........     193

940 CACCGAGGCCTCCGATGGGGGGAAGGGCTGGGCGGCACCGACCTGCGGGG    989
    |||||||||||   ||| :::::|||
194 ..ProArgProAlaTrpArg.AlaAlaAlaArgGlyProAlaSerGl     209

990 GCCG...GACTGGGCCACTGCCGTAGCACTCGCCTGCTGGAGCACCGA    1036
    |||  |||||                        ::||||||
209 yProThrGlyLeuGly.....................SerGlyAlaGlyT  219

1037 TGCGCCCGTGCCTTCGACGAGTGGAACTGACAGCCGCCAAGGCTGATTG    1086
     ||||||                |||    ||| :::::::::||
219 rpArgArgProArgSerSerGlyArg.......................  228

1087 CTGGCTGCGGCCCAGCACTTGCCTGACGGCCTTGACCTGCCGCCCTCA    1136
     ||||||  ||||||     |||    ||| |||:::::::||
229 ...AlaTrpGlyProArgProValProGlyProTrpProAlaGluProAl  244

1137 AGGCCGC............AGCCCAGGGCT......CTTCCTGCTA    1165
     |||                :::||::|||           :::
244 aArgArgThrArgSerAsnValThrProGluAlaAlaTrpValPheArgG  261

1166 CTTGCGCCACTGGACACAAAACCTGCAGCTACACCTGCTGTGCTACAGCCC    1215
     ||||||  :::      ||||  :::|||    :::|||
261 lyAlaProGlySerSerAlaProSerArgSerProGluArgArgTrpGln   277

1216 AGCGAACGTGTGAAGGCTGCCCCCTGCTTGGGCTGGGCCCCCACCCA    1265
     :::  |||:::         |||||||||||:::|||   :::
278 GluProArgIleTyrThr............LeuGlyAlaSerProProSe   290

1266 ACACACTCAAGTCACTGCGCCCCAGGCTGGCCTCTTGGTGCTGGGAAAG    1315
     :::      :::::  |||::::||::||::::||  ||:::
290 rGlnGlyGlyProProArgGlyArgAspTrpProIleGly...GlyArg   306

1316 TGTAGGCTGGTGCCAGCCTGTCCCCCACTGCTTCTTACTCCCTCCCTGGA    1365
     |||:::||          ||::::::::     ::: |||
306 lnArgGlyGlyAla.......ArgProThrSerValArgArgValPheGly  320

1366 GCCCTC...........TTGCCCCCA    1380
     |||               ||||||||
321 AlaArgProIleGlyIleArgGluLeuProPro   330

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-231-193A-58

seq_documentation_block:
; Sequence 58, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie P.
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen W.
;    APPLICANT:  Lu, Chin-Chun
;    TITLE OF INVENTION:  HUMAN N-METHYL-D-ASPARTATE RECEPTOR
;    TITLE OF INVENTION:  SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Brown, Martin, Haller & McClain
;       STREET:  1660 Union Street
;       CITY:  San Diego
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
```

```
         OPERATING SYSTEM: PC-DOS/MS-DOS
         SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
         APPLICATION NUMBER: US/08/231.193A
         FILING DATE: 20-APR-1994
         CLASSIFICATION: 536
     PRIOR APPLICATION DATA:
         APPLICATION NUMBER: US 08/052,459
         FILING DATE: 20-APR-1993
         CLASSIFICATION: 536
     ATTORNEY/AGENT INFORMATION:
         NAME: Seidman, Stephanie
         REGISTRATION NUMBER: 33,779
         REFERENCE/DOCKET NUMBER: 6362-9383
     TELECOMMUNICATION INFORMATION:
         TELEPHONE: 619-238-0999
         TELEFAX: 619-238-0062
 INFORMATION FOR SEQ ID NO:  58:
     SEQUENCE CHARACTERISTICS:
         LENGTH: 1336 amino acids
         TYPE: amino acid
         TOPOLOGY: linear
     MOLECULE TYPE: protein
 08-231-193A-58

alignment_scores:
       Quality:  191.00          Length:     450
       Ratio:      1.194         Gaps:        24
Percent Similarity: 35.556     Percent Identity: 24.667

alignment_block:
US-09-786-136-4 x US-08-231-193A-58    .

Align seg 1/1   to: US-08-231-193A-58    from: 1  to: 1336
```

seq_documentation_block:
; Sequence 58, Application US/08486273A
; Patent No. 5985586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTRO SUBUNITS, DNA
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,273A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-486-273A-58

nment_scores:
Quality: 191.00   Length: 450
Ratio: 1.194   Gaps: 24
Percent Similarity: 35.556   Percent Identity: 24.667

alignment_block:
US-09-786-136-4 x US-08-486-273A-58   :

Align seg 1/1 to: US-08-486-273A-58   from: 1 to: 1336

353 GGGGCCCTGCAGTGTGCAGCTCAGAGCCCCTGAGCCCCCAGGAACCCC 402
     ||| ::: ||||||||:::||| || ||||||::: ||
892 GlyMetTyrSerCysSerAlaGluAlaAlaProProProAlaLysPr 908

403 TCCTGCCTCTCAGCCATCTAGATGCAGCTCCTCTGCCCACTGTTGTCT 452
     ||| ||||||||||:::
908 oProPro.........ProProGlnProLeuProSer....... 917

453 ACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGCGCTGGGACTCGGAC 502
917 917

503 CAAAATGGCAACTCCAAGCCGTGCTTTGGGGACCCTGCCACTCCCACGA 552
     ||||| |||:::
918 ....ProAlaTyrProAlaPr 923

553 AGGTCTTCGCGCGCCACCT..................CCCGTCCTCCCTGTCGGC 590
     ||||||| ||||||
923 oGlyProAlaProGlyProAlaProPheValProArgGluArgGlnArgAlaSerV 940

591 TCAGCATGGGCCCGTCACAAACTCGTAGCCTGACCCGGCCCAGGCC 640
     ::::: |||||| ||| ||||||||| ||||| ||||
940 aAlaArgTrpArgArgProLysGlyAlaGlyPro...ProGlyGlyAla 955

641 AACAACAGTGAAGGCAGCGACCATGACTACCTGCCCTG........ 679
     ::: ::::|:||| ||| |||:::
956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnGl 972

680 ......GTGGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACGCGC 722
       ||| |||||||| :::||| |||
972 yLeuGlyLeuGlyLeuGlyGluAla.ArgAla............ 982

723 CCTTCTGCGCCGCTGGTGCCATCTGCCAGGAGCGCCTCTGCGTGCCTCG 772
     ||||| |||::: |||
983 ......AlaProArgGlyAlaAlaGlyArg........... 990

773 CCCTTTGCCGTGCACCCGCCAGCCTCAGCCCCACCTCGGCCTCATTGCC 822
     ||| |||||::: ||||||:::
991 .....ProLeuSerProProAlaAlaGlnProProGlnLysPro...P 1004

823 CTGGGCACTTCTGGGCCCTGTGTTGGCCAGGTGACCAGTG......... 863
     ||:::::|||::::|||:::
1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGlaProPro 1020

864 ..............CCACGGCCTCCTGCAGCCCGTCCCCCAGC 892
     ....ProProAl 1035
1021 AlaGlyAlaPheProGlyPheProSerProProAla.....

893 TCGGGCTCTGAGG.................... 905
     ||| |||
1035 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052

906 ...............GGCCAGGCCAGTGGACA....... 923
     ::: ||||||||||
1052 spGluSerProProAlaProAlaArgTrpProArgSerAspProGluSer 1068

924 ...............GTGGCGGGGCTCAGACACCGAGGCCTCC 952
     ::||||||||||
1069 GlnProLeuLeuGlyProGlyAlaGlyGlyAlaGlyGlyThrGlyGlyAl 1085

953 GATGGGGCGGAAGGCGTGGGCGGCCACCGACCTGCGGGGCCGACCTGGGC 1002
     ||| ||| ||| ||||||:::::||||
1085 aGlyGlyGlyAlaProAlaProAlaProProCysPheAlaAlaProProP 1102

1003 CACTG..........CCGTAGCACTCGCCTGGCTGGAGCACCGATGCG 1040
     :::||||||::: ||| |||:::
1102 roCysPheTyrLeuAspValAspGlnSerProSerAspSerGluAspSer 1118

1041 CCGCTGCCTTCGACGAGTGGGAACTGACAGCGGCCAAGGCTGATTGCTGG 1090
     ||| ||||||| ||||||
1119 .....GluSerLeuAlaGly.................AlaSerLeuAlaGl 1128

1091 CTGCGGGGCCCAGCACTTGCCTGCGACGGCTTGACCTGCCGCCCTCAAGGC 1140
     || ||| |||
1128 yLeuAspPro...................TrpTrpPheAlaAspPheP 1138

1141 CGCAGCCC.......... 1148
     |||
1138 roTyrProTyrAlaAspArgLeuGly**ProAlaAlaArgTyrGlyLeu 1154

1149 .......................GAGGGCTCTTCCTGCTACTGCGCCAC 1174
     ||:||| :::
1155 ValAspLysLeuGlyTrpLeuAlaGlySerTrpAspTyrLeuPro** 1171

1175 TGGGACCAAAACCTGCAGCTACACCTGC.....TGTGCTCAGCCCCAGC 1218
     ::: ::::::|| ||| ||||||:::
1171 *ArgSerGlyArgAlaAlaTrpHisCysArgHisCysAlaSer....... 1185

1219 GAACGTGTGAAGGCTGAAGGCTGCTTGGGCTGGCGCCCCACCCAACA 1268

Percent Similarity: 35.556   Percent Identity: 24.667

alignment_block:
US-09-786-136-4 x US-08-940-086A-58   .

Align seg 1/1   to: US-08-940-086A-58   from: 1 to: 1336

```
  353 GGGGCCCAGGTGTGCCAGCTCAGTTCAGTGCCCGCTGAGCCCCAGGAACCCC  402
      ||      :::  ||||||||::  ||||||:|:  |||||| ||
  892 GlyMetTyrSerCysSerAlaGluAlaAlaProProAlaLysPr  908

  403 TCCTGCCTCTCACAGCCATCTAGATGCAGTCCTCTGCCCACTGTTGTCT  452
      ||||        ||  ||||||||||| |||||||||:::
  908 oProPro.........ProProGlnProLeuProSer  917

  453 ACTCTAAAGGACTTCAGAGAGGCTCTCCAGCAGGCGCCTGGGACTCGGAC  502
      ||||||||||||| ||  :|||||
  917                                                    917

  503 CAAAATGCAACTCCAAGCGTGCTTGGGGACCCTGCCACTCCCACGGA  552
                                        |||||| |||::
  918 ........................ProAlaTyrProAlaPr  923

  553 AGGTCCTCGCCGCCACCT.........CCCCGTGCTCCCTGTCGGC  590
      ||||||         ||              |||||   :::
  923 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaSerV  940

  591 TCAGCATGGGCGCCGCCGTCACAAACTCTGTAGCCCGGACCCGGGCCAGGCC  640
      ||:::      |||||   |||   :::::||  |||||| |||
  940 aIaAlaArgTrpArgAlaProLysGlyAlaGlyPro...ProGlyGlyAla  955

  641 AACAACAGTGAAGGCAGCGACCATGACTACTACCTGCCCTG.........  679
      :::  :::::|||   |||  :::::|||
  956 GlyLeuAlaAspGlyPheHisArgTyrTyrGlyProIleGluProGlnGl  972

  680 .......GTGCGGCTGCAGGAGGCACCAGGCTCCTTCGCGCCTGGACGCGC  722
            :::  ||| |||||| :::|||
  972 yLeuGlyLeuGlyLeuGlyGluAla.ArgAla.................  982

  723 CCTTCTGCGGCCGCTGTGGCCATCTCGGCAGGAGCGCCTCTGCCGCTGCCTCG  772
      ||||| |||||:::||  |||
  983 .......AlaProArgGlyAlaAlaGlyArg.................  990

  773 CCCTTGCCGTGCCACCGCCGCCAGCCTCAGCCCCACCTCGGCCTCATTGCC  822
      |||  :::  |||||||  |||||||  :::|||
  991 .......ProLeuSerProProAlaAlaGlnProProGlnLysPro....P 1004

  823 CTGGGCACTTCTGGCCCTGGTTGGCCAGGTGACAGTG.........  863
      ||:::  :::::|||:::|||     :::
 1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGluProPro 1020

  864 ........................CCACGGCCTCCTGCAGCCCGTCCCCCAGC  892
                              |||||||  ||||||||
 1021 AlaGlyAlaPheProGlyPheProSerProAla......ProProAl 1035

  893 TCGGGCTCTGAGG..............................  905
      |  |||||
 1035 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052

  906 ...............GGCCAGGCCAGGTGGACA...............  923
                     :::  |||||||||
 1052 spGluSerProProAlaProAlaArgTrpProArgSerAspProGluSer 1068

  924 ....................GTGGGCGGGGCTCAGCACACCGAGGCCTCC  952
                          ::|||||||
 1069 GlnProLeuGluGlyProGlyAlaGlyGlyAlaGlyGlyThrGlyGlyAl 1085

  953 GATGGGCGAAGGGCTGGGCGGCACCGACCTGCGGGGCCGGGACCTGGGC 1002
      |||            :::|||   |||||||||
 1085 aGlyGlyGlyAlaProAlaAlaProProProCysPheAlaAlaProProP 1102

 1003 CACTG.........CCGTAGCACTCGCCTGGCTGGACACCGATGCG 1040
      ||               :::||||||:::   |||  |||:::
```

1186 ....LeuGluLeuLeuProProProArg 1193

1269 CACTCAAGTCACTGCCGCCCAGGGCTGCCCTCCTTGGTGTCGTGGGAAAGTGT 1318
     || ||| ||::: ||||| :::
1194 HisLeuSerCysSerHisAspGlyLeuAspGly............. 1204

1319 AGGCTGGTGCGCAGCCTGTCCCCCACTGCTTCTTCTACTCCTCCTCCTCCGGAGC. 1367
     |||||| ||| |||||||||
1205 .GlyTrpTrpAlaProProPro...........ProProTrpAlaA 1216

1368 ....CCTCTTGCCCCCACAAAAGTGCCTGCCTGTGC......TCTCTC 1406
     |||||| ::::: ||| || |||
1216 laGlyProLeuProArgArgArgAlaArgCysGlyCysProArgSerHis 1232

1407 CCTCTCCTCCCACCCCACTCACACTCCCCTCCATCCTCTGAGCTCCCT 1454
     |||||| |||||| |||:::::: |||
1233 ProHisArgProArgAlaSerHisArgThrProAlaAlaAlaAlaPro 1248

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-940-086A-58

alignment_scores:
  Quality: 191.00   Length: 450
  Ratio: 1.194   Gaps: 24

```
        PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/052,449
;       FILING DATE:  20-APR-1993
;       CLASSIFICATION:  536
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6362-9383E
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-238-0999
;       TELEFAX:  619-238-0062
;   INFORMATION FOR SEQ ID NO:  58:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH:  1336 amino acids
;           TYPE:  amino acid
;           TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
;   US-08-940-035A-58

alignment_scores:
    Quality:   191.00            Length:         450
    Ratio:       1.194           Gaps:            24
Percent Similarity:  35.556      Percent Identity:  24.667

alignment_block:
US-09-786-136-4 x US-08-940-035A-58

Align seg 1/1  to: US-08-940-035A-58   from: 1  to: 1336

 353 GGGGCCCTGCAGGTGTGCAGCTCAGAGCCCGCTGAGCCCCAGGAACCCC  402
     |||   |||   ::: |||||||:::||| |||  ||||||||:::  ||
 892 GlyMetTyrSerCysSerAlaGluAlaAlaProAlaProAlaLysPr    908

 403 TCCTGCCTCTCACAGCCATCTAGATGCAGCTCTGCCCACTGTGTCT      452
     |||                    |||||||||||||||||||||
 908 oProPro......................ProProGlnProLeuProSer   917

 453 ACTCTAAAGACTTCAGAGAGCTCTCCAGCAGGCGGCCTGGGACTCGGAC   502
                                  ||||||||||:
 917 .................................................   917

 503 CAAAATGGCAACTCCAAGCGTGCTTTGGGGACCCGCACTCCCACGGA     552
                                               ||||| |||::::
 918 ..................................ProAlaTyrProAlaPr  923

 553 AGGTCCTGCCGCCCCACCT................CCCCGTTCCTCCGTCTGGGC  590
     |||||                              ||  :: ::|
 923 oGlyProAlaProGlyProAlaProPheValProArgGluArgAlaSerV  940

 591 TCAGCATGGGCGCCGTCACAAACTCTGTAGCCTCGACCCGGCCAGGCC     640
     ::|:: |||||| ||| |||:::::||| ||||||||| |||||||
 940 aIAlaArgTrpArgArgAlaArgCysGlyAlaGlyLyPro...ProGlyGlyAla  955

 641 AACAACAGTGAAGGCAGGCAGCCACCATGACTACCTGCCCTTG........  679
               ::::::|||   || ||| ||||||| |||:::
 956 GlyLeuLeuAspGlyPheHisArgTyrTyrGlyLyProIleGluProGlnGl  972

 680 .......GTGCGGCTGCAGGAGGCACCAGGCTCCTTCCGCCTGGACGCGC  722
                 ||  ||||| ||||||||||:::|||
 972 yLeuGlyLeuGlyLeuGluAla.ArgAla.....................  982

 723 CCTTCTGCCGCCGTGTGCGCCATCTCCGCCAGGAGGCCCTCTGCGTGCCTCG  772
     ||||||           |:::|||||| |||
 983 .....AlaProArgGlyAlaAlaGlyArg.....................  990

 773 CCCTTTGCCGTGCCACCCGGCACCTCAGCCCCACTTCGGCCTCATTGCC    822
     |||  ::: |||||||||  |||||::: |||
 991 .....ProLeuSerProProAlaAlaGlnProProGlnLysPro...P   1004

 823 CTGGGCACTTCTGGCCTGGTGTTGGCCAGGTGACAGTG.............  863
     ||::::|::||||::|||||||
1004 roAlaSerTyrPheAlaIleValArgAspLysGluProAlaGluPro   1020
```

```
1102 roCysPheTyrLeuAspValAspGlnSerProSerAspSerGluAspSer 1118
1041 CCGCTGCCTCGACGAGTGGGAACTGACAGCCGGCCAAGGCTGATTGCTGG 1090
                 |||    :::||||
1119 .....GluSerLeuAlaGly....................AlaSerLeuAlaGly  1128
1091 CTGCGGGCCCAGCACTTGCCTGACGGCCTTGACCTGCCGCCCTCAAGGC  1140
     |      |||                   |||
1128 yLeuAspPro...................TrpTrpPheAlaAspPheP  1138
1141 CGCAGCCC....................................... 1148
     ||
1138 roTyrProTyrAlaAspArgLeuGly***ProAlaAlaArgTyrGlyLeu  1154
1149 ...............................GAGGGCTCTTCCGCTACTGCGCCAC 1174
                                    |||||   ||  ||
1155 ValAspLysLeuGlyGlyTrpLeuAlaGlySerTrpAspTyrLeuPro**  1171
1175 TGGGACCAAAACCTGCAGCTACACCTGC.......TGTGCTACAGCCCAGC 1185
      : |:::|||  :::||||||||||:: |||
1171 *ArgSerGlyAlaAlaIleTrpHisCysArgHisCysAlaSer........ 1171
1219 GAACGTGTGAAGGCTGCCCCCTGCCTGCTGGGCTGGCGCCCACCCAACA 1268
     |||              |||  ||| ||| |||
1186 ...............LeuGluLeuProProProArg..ProProTrpProAlaA 1193
1269 CACTCAAGTCACTGCCCGCCCAGGCTGCCTCGCCTGGGTGCTGGGAAAGTGT 1318
     |||   |||        |||||
1205 .GlyTrpTrpAlaProProPro.....................ProProTrpPro 1216
1319 AGGCTGGTGCCAGCCTGTCCCCCCACTGCTTCTTACTCCCTCCTGGAGC. 1367
     ||||||   |||             ||||||:::
1368 ....CCTCTTGCCCCCACAAAAGTGCTGCCTGTGC.....TCTCTC  1406
        |||||         ::: |||  ||| |||
1216 laGlyProLeuProArgArgArgArgCysGlyCysProArgSerHis    1232
1407 CCTCTCCTCCCACCCCACTCACACTCCCTCCCACATCCCTCAGCTCCCT  1454
     ||      |||||        ||||||  |||
1233 ProHisArgProArgAlaSerHisArgThrProAlaAlaAlaAlaPro   1248

seq_name:  /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-940-035A-58

seq_documentation_block:
;  Sequence 58, Application US/08940035A
;tent NO. 6316611
ENERAL INFORMATION:
;   APPLICANT:  Daggett, Lorrie P.
;   APPLICANT:  Ellis, Steven B.
;   APPLICANT:  Liaw, Chen W.
;   APPLICANT:  Lu, Chin-Chun
;   TITLE OF INVENTION:  HUMAN N-METHYL-D-ASPARTATE RECEPTOR
;   TITLE OF INVENTION:  SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
;   NUMBER OF SEQUENCES:  63
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Heller Ehrman White & McAuliffe
;       STREET:  4250 Executive Square, 7th Floor
;       CITY:  La Jolla
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92037
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/940,035A
;       FILING DATE:  29-SEPT-97
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/231,193
;       FILING DATE:  20-APR-1994
```

APPLICANT: Goldstein, Steven A.
APPLICANT: Lin, Wushan
TITLE OF INVENTION: METHODS AND COMPOSITIONS
TITLE OF INVENTION: FOR STIMULATING BONE CELLS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 30-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1442 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-316-650-12

alignment_scores:
Quality: 189.50   Length: 566
Ratio: 0.746   Gaps: 34
Percent Similarity: 44.876   Percent Identity: 26.325

alignment_block:
US-09-786-136-4/rev x US-08-316-650-12

Align seg 1/1 to: US-08-316-650-12 from: 1 to: 1442

```
1455 CAGGGAGCTCAGAGGATGGAGGGGAGTGTGAGTGGGGTGGGAGGAGAGGG 1406
     :::|||||||||      :::||||||  ::::  |||::: |||:::: |
 341 GluGlyAlaGlnGlySerArgGlyGluProGlyAsnProGlySerProGl 357

1405 AGAGACACAGGCAGGCACTTTTGTGGGGCAAGAGGGCTCCAGGGAGG 1356
     |  ||| |||||::||     |||||||||  ||||||
 357 yPro.AlaGlyAlaSerGlyAsnProGlyThrAspGlyIleProGlyAla 373

1355 GAGTAAGAAGCAGTGGGGACAGCTGCCACCAGCCTACACTTTCCCAGC 1306
     ::: :||||  ::::|||    :::||| ||| :::||||||:::
 374 LysGlySerAlaGlyAlaProGlyIleAlaGlyAlaProGlyPheProGl 390

1305 ACCAAGAGGCCAGCCTGGGCGCCAGTGACTTGAGTGTGTGGTGGGGC 1256
     ||||||||| |||  ::|||||||||  ||||||  ||| |||||
 390 yProArgGlyProPro.................GlyP 397

1255 GCCAG................CCCAAGCAGCAGGGG 1236
     |||  ||||||    |||:::
 397 yProArgGlyProPro.

 397 roGlnGlyAlaThrGlyProLeuGlyProLysGlyGlnAlaGlyGluPro 413
1235 .....GCAGCCTTCACACGTTCGCTGGGC.........TGTAGCAC 1204
```

```
 864 ................CCACGGCCTCCTGCAGCCCGTCCCCAGC 892
                    |||  ||||||||
1021 AlaGlyAlaPheProGlyPheProSerProProAla......ProProAl 1035

 893 TCGGGCTCTGAGG 905
     |  |||
1036 aAlaAlaAlaThrAlaValGlyProProLeuCysArgLeuAlaPheGluA 1052

 906 ...........GGCCAGGCCAGGTGGACA 923
                ::: |||||||
1052 spGluSerProProAlaProAlaArgTrpProArgSerProGluSer 1068

 924 ........GTGGGGCGGGCTCAGACACCGAGGGCCTCC 952
             ::|||||||
1069 GlnProLeuLeuGlyProGlyAlaGlyGlyAlaGlyThrGlyGlyGlyAl 1085

 953 GATGGGGGCGAAGGGCTGGGCGGCCACCGACCTGCGGGGCCGGACCTGGGC 1002
     |||   ||| |||   ||| |||::::|||  |||:::|||
1085 aGlyGlyGlyAlaProAlaAlaAlaProProProProCysPheAlaAlaProProP 1102

1003 CACTG.......CCGTAGCACTCGCCTGGCTGAGCACCGATGCG 1040
     ||                :::||||||  :::|||  |||:::||
1102 roCysPheTyrLeuAspValAspGlnSerProSerAspSerGluAspSer 1118

1041 CCGCTGCCTCGACGAGTGGGAACTGCAGCGGCCAAGGCTGATTGCTGG 1090
                                       ::||||||
1119 .....GluSerLeuAlaGly...........AlaSerLeuAlaGl 1128

1091 CTGCGGGGCCCAGCCACTTGCCTGACGGGCGCCTTGGCCGCCTCAAGGC 1140
                                       ||| ::: |||
1128 ...........TrpTrpPheAlaAspPheP 1138

1141 CGCAGCCC.......... 1148
     ||  |||
1138 yLeuAspPPro...............

1149 ...........GAGGGCTCTTCCTGCTACTGCGGCCAC 1174
                |||||
1138 roTyrProTyrArgAspArgArgLeuGly***ProAlaAlaArgTyrGlyLeu 1154

1155 ValAspLysLeuGlyGlyTrpLeuAlaGlySerTrpAspTyrAspTyrLeuPro** 1171

1175 TGGGACCAAAACCTGCAGCTACACCTGC.....TGTGCTACAGCCCAGC 1218
     :  ::::||||     ::::||||||:::
1171 *ArgSerGlyArgAlaAlaTrpHisCysArgHisCysAlaSer....... 1185

1219 GAACGTGTGAAGGCTGCCCCCTGCTGTGGGCTGGCCGCCCACCCAACA 1268
     |||   ||| |||    :::|||||||||||||:::
1185 .............ProProTrpPAlaA 1216

1186 ............LeuGluLeuLeuProProProArg 1193

1269 CACTCAAGTCACTGCCGCCCCAGGCTGCCCTTGGTGCTGGGAAAGTGT 1318
     |||||  |  :::|||
1194 HisLeuSerCysSerHisAspGlyLeuAspGly................. 1204

  19 AGGCTGGTCCCAGCCGTGTCCCCACTGGTTCTTACTCCCCTCCTGGAGC. 1367
     |||||   |||   |||||  :::|||
1205 .GlyTrpTrpPrpAlaProPro.......ProProTrpPAlaA 1216

1368 ....CCTCTTGCCCCCACAAAAAGTGCCTGCCTGTGC......TCTCTC 1406
         ::::||| |||  ||| |||
1216 laGlyProLeuProArgArgArgAlaArgCysGlyCysProArgSerHis 1232

1407 CCTCTCCTCCCACCCCCACTCACCTCCCCCATCCCTCGAGCTCCCT 1454
     |||   |||||||     ||||||
1233 ProHisArgProArgAlaSerHisArgThrProAlaLeuAlaAlaPro 1248
```

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-316-650-12

```
414 GlyIleAlaGlyPheLysGlyAspGlnArgGlyProLysGlyGluThrGlyPr 430
        :::|||                              :::
1203 AGCAGGTGTAGCTGCAGGTTTGGTCCCAGTGGC..............G 1169
        :::::|||     |||||::|||
430 oAlaGlyProGlnGlyAlaProGlyProAlaGlyGluGluGlyLysArgG 447

1168 CAGTAGCAGGAGGAGCCCTGCGGGCTGCGGCCTTGAGGGCGGCCAGGTCAA 1119
        :::::  |||::::              |||   ||||||::||
447 lyAlaArgGlyGluProGlyGlyAlaGlyProIleGlyProProGlyGlu 463

1118 ...................GGCCGTCAGGCAAGTCCTGG 1099
                               ||||||::||     ||||
464 ArgGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGl 480

098 GCCC.......................................... 1095
480 yProLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProL 497

1094 ..GCAGCCAGCAATCAGCCTTGGCCGCTGTCAGTTCCCACTCGTCGAAGG 1047
        ::::|||::::::||| |||||           |||   :: ||
497 ysGlyAlaAsnGlyAspPro.GlyArgProGlyGluProGlyLeuProGl 513

1046 CAGCGGCGCATCGGTCTCCAGCCAGGCGAGTGCTACGGCAGTGGCCCAG 997
        |::: ::  ::: :::||||||::::              |
513 yAlaArgGlyLeuThrGlyArgProGlyAspAla...............G 525

996 GTCCGGCCCCGCAGGTCGGTGCCGCCCAGCCCTTCCGCCCATCCGGAGGC 947
        |||||  |||||||||           ||| |||||::::
525 lyProGlnGlyLysValGly...............ProSerGlyAla 535

946 CTCGGTGTCTGAGCCCCGCCACTGTCCACCTGGCCTGGCCCCTCAGAGC 897
        ||| :||||| |||                   |
536 ProGly.GluAspGlyArg...............ProGlyProProGlyP 547

896 CCGAGCTGGGGACGGGCTGCGAGGAGCCGTGGCACTGTCACCCTGGCCA 847
        |||::: ::: |||    :::::::::
547 roGlnGlyAlaArgGlyLeuGlnProGlyValMetGlyPhe.......... 559

846 ACACCAGGGCCCAGAAGTCCCAGGCAATGAGGCCCGAGGTGGGCTGAG 797
        ||||||||||::::|||::::||:::|||     ||||||
560 ...ProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly...... 573

796 GCTGCCGCGGTGCACGGCAAAGGGCAAAGGGCACGGC..........AGA 759
                                   |||  |||     ||
574 ..............GluLysGlyLeuAla.GlyAlaProGlyLeuArg 584

758 GGCGCTCCTGCGAGATGCGACACGCGCCCAGAGAGGGCGCGTCCAGGCGG 709
        ||| |||: :::::              ::  ||||||
585 GlyLeuProGlyLysAspGlyGluThrGlyAlaAlaGlyProProGly.. 600

708 AAGGAGCCTGGTGCTTCCTGCAGCCGCCACCAAGGGCCAGGTAGTCATGGTC 659
        ||| ||||||::||||||                 :::||
601 .......ProSerGlyProAlaGlyProGlyArgGlyGluGlyGlnGlyA 613

658 GCTGCCCTCACTGTTGTGTCCGGGTCGGGTCAGGGTCACAGAGTTGT 609
        |||::: ::: |||||||||::::::|||:::|||
613 la...............ProGlyPro..SerGlyPheGlnGlyLeuP 623

608 GACGGGCGGCCCATGCTGAGCCCACAGGAGGACGGGGAGGTGGGCGCGA 559
        |||              ::||||||||: |||  :::
623 roGlyProProGlyProProGlyGluGlyLysGlnGlyAspGlyGly 639

558 GGACCTTCCGTGGGAGTGGCAGGGTCCCC........AAAGCACGCTT 518
        |||   ||||||||||||                  |||
640 IlePro.....GlyGluAlaGlyAlaProGlyLeuValGlyProArgG 654

517 GGAGTTGCCATTTTGGTCCGAGTCCCAGCGCCTGCTGGAGAGCCTCTCT 468
        |||   :::||::::::::::              |||
654 ygluArgGlyPheProGlyGluArgGlySerProGlyAlaGlnGlyLeuG 671

467 GAAGTCCTTAGAGTAGAACAACAGTGGGCAGAGGAGCTGCATCTAGATGG 418
        :::|||
```

```
671 lnGlyPro......................ArgGlyLeuPro.......... 677
                                                  ||||||::|||
417 CTGTGAGAGGCAGGAGGGGTTCCTGG........GGCTCAGC 383
                                    |||::|||||||
678 ............GlyThrProGlyThrAspGlyProLysGlyProLysGlyAlaAl 689

382 GGGCTCTGAGCTGCACACCTGCAGGGCCCCAGGCAGGACTCCCTAGTAG 333
        :::         |||   |||||  |||||||
689 aGlyProAsp..Gly.ProProGlyAlaGln...GlyPro.......... 700

332 TAGCATCCACAGTCACAGGCCAAGTGAAGCAGGAGGGGGCAGAGCTGACC 283
        |||  |||||||||     |||||::||| ||| |||
701 .....ProGlyLeuGlnGlyMetProGlyGluArg.GlyAlaAlaGlyI 715

282 TTACTTGTCTGAAGGGCTCGAAGCCA................. 257
        ::::  |||||||::::::
715 leAlaGlyProLysGlyAspArgGlyAspValGlyGluLysGlyProGlu 731

256 ..............................................GCACC 252
                                                      |||
732 GlyAlaProGlyLysGlyGlyArgGlyArgLeuThrGlyProIleGlyPr 748

251 TCCGGGCATGGCCTGTGGAACCAAGGAGCAGTCTGTGGAGCCCCCACCA 203
        ||||||| |||      |||||||::::||         |||||
748 oProGlyProAlaGlyProAlaAlaAsnGlyGluLysGlyGluAlaGlyLysProProG 765

202 ............CCCACCCTCTACAGAGGGCTCATAGAAAAAATTGCAGTG 164
                             |||                :::
765 lyProSerGlySerThrGlyAlaArgGlyAlaProGlyProGlyGlu 781

163 TTCTCCCAGCTTCTCCCTCTATCGACTTTAACAGAGTAGATTGCCTTGC 114
        ::|||       ::::|||
782 ThrGlyProProGlyProAla......................GlyPheAl 791

113 TAGGCCCAGAACCAGCTCAGACTCCCCAACTTTGACAGAGAAGTGATCAGG 64
        |||||| |||      :::|||        ::::::||||||||
791 aGlyProProGlyAlaAlaAspGlyGlnProGly...AlaLysGlyAspGlnG 807

63 CTGAATCTTCAACCCATAGCAGAGCAAGAGCCCCC 29
        ::|||:::::     ::: |||    ||||||
807 lyGluAlaGlyGlnLysGlyGlyAlaGluAspGlyAlaGlyAlaPro 818
```

```
;       FILING DATE:   18-FEB-1994
;       CLASSIFICATION:
;       ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UMIC009P--
;       TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (713) 789-2679
;       TELEX:  79-0924
;   INFORMATION FOR SEQ ID NO:  12:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH:  1442 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
PCT-US95-02251-12
```

```
ment_scores:
        Quality:    189.50        Length:         566
        Ratio:      0.746         Gaps:            34
tent Similarity:    44.876        Percent Identity:  26.325
```

```
alignment_block
US-09-786-136-4/rev x PCT-US95-02251-12   ..

Algn seg 1/1   to: PCT-US95-02251-12   from: 1   to: 1442
```

```
1455  CAGGGAGCTCAGAGGATGGACGGGAGTGAGGGGAGTGGGGTGGGAGGAGAGGG 1406
      ::||||||||            :::||||||||    :::::  |||::::    ||
 341  GluGlyAlaGlnGlySerArgGlyArgGlyAsnProGlySerProGl         357

1405  AGAGAGCACAGGCAGGCAGCACTTTTGTGGGGCAAGAGAGGGCTCCAGGGAGG 1356
      |  || |||:::       |||       ||||||  |||||||||
 357  yPro.AlaGlyAlaSerGlyAsnProGlyThrAspGlyIleProGlyAla    373

1355  GAGTAAGAAGCAGTGGGGACCAGGCTGGCCACCAGCCTACACTTTCCCAGC  1306
      :::  ::||| |||  :::||| |||  :::||| |||  ||||||:
 374  LysGlySerAlaGlyAlaProGlyIleAlaGlyAlaProGlyPheProGl   390

1305  ACCAAGAGCCCAGCCCTGGGCGCGCAGTGACTTGAGTGTGTTGGGTGGGGC  1256
      :||||||||  |||               ||||||| ||| |||::|||
 390  yProArgGlyProPro.................................GlyP 397

1255  GCCAG.................CCCAAGCAGCAGGGG............ 1236
      |||                   |||||| |||:::
 397  roGlnGlyAlaThrGlyProLeuGlyProLysGlyProAlaGlyGluPro  413

1235  ......GCAGCCTTCACACGTTCGCTGGGC..........TGTAGCAC 1204
      ||:::||| |||  |||                        :::
  14  GlyIleAlaGlyPheLysGlyAspGlnGlyProLysGlyGluThrGlyPr 430

  03  AGCAGGTGAGCTGACGTTTGGTCCCAGTGGC................G 1169
      ||||||||  |||:::  ||||||:::||
 430  oAlaGlyProGlnGlyAlaProGlyGluGluGlyLysArgG         447

1168  CAGTAGCAGGAAGAGCCCCTCGGGCTGCGGCCTTGAGGGCGGCCAGGTCAA 1119
      :::::  ||||:   |||       ||| |||  ||||||:::
 447  lyAlaArgGlyGluProGlyLysProAlaGlyGlyProIleGlyProGlyGlu 463

1118  ...................GGCCGTCAGGCAAGTGCTGG 1099
      |||:::|||| |||||
 464  ArgGlyAlaProGlyAsnArgGlyPhePheProGlyGlnAspGlyLeuAlaGl 480

1098  GCCC............................................ 1095
      ||||
 480  yProLysGlyAlaProGlyArgGlyProSerGlyLeuAlaGlyProL  497

1094  ..GCAGCCAGCAATCAGCCTTGCCGCTGTCAGTTCCCACTCGTCGAAGG 1047
      :::||||||:::||| ||||||  ||| :::
 497  ysGlyAlaAsnGlyAspPro.GlyArgProGlyGluProGlyLeuProGl 513
```

(right column)

```
1046  CAGGCGGGCCATCGGTCGGTGCTGCCAGCCAGGCGAGTGCTACGGCAGCGAGTGGCCCAG 997
      |::.:::  ||| .:.    :::|||||:::|||
 513  yAlaArgGlyLeuThrGlyArgProGlyAspAla.......................G 525

 996  GTCCGGCCCGCAGGTCGGTGCGGCCCAGCCCCTTCCGCCCATCGGAGGC 947
      |||||      ::.|||||||                        |||  ||:::|
 525  lyProGlnGlyLysValGly.....................ProSerGlyAla 535

 946  CTCGGTGCTGAGCCCGCCCACTGTCCACCTGGCCTGGCCCCTCAGAGC 897
      |||  :::::|||  |||::::||                 |||||||||
 536  ProGly.GluAspGlyArg......................ProGlyProProGlyP 547

 896  CCCAGCTGGGGACGGCGTGCAGGAGGCCGTGGCCACTGTCACCCTGGCCA 847
      |::.::     |||   |||::::|||||:::||         ::::|||:::
 547  roGlnGlyAlaArgGlyGlnProGlyValMetGlyPhe.............. 559

 846  ACACCAGGCCCAGAAGTGCCCAGGGCAATGAGGCCCAGGTGGGCTGAG 797
      |||||||||:::||:::||||::  :::|||||||:::
 560  ...ProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly....... 573

 796  GCTGGCGCGGTGCACGGCCAAAGGGCAGGCCACGGC............AGA 759
      ||||||  ||| ||| |||                          |||
 574  ..............GluLysGlyLeuAla.GlyAlaProGlyLeuArg 584

 758  GGCGGCCTCCTGCGAGATGCCGCACAGCGGCCGCAGAAGGGCGTCCAGGCGCGG 709
      ||| ||:::|||    |||:::|||||||:::   ||      :::||||||
 585  GlyLeuProGlyLysAspGlyGluThrGlyAlaIleAlaGlyProProGly.. 600

 708  AAGGAGGCCTGGTCCCCTGCAGCCGCACCAAGGGCAGCAGGTAGTCATGGTC 659
      |||||||||          ::.::    ||| |||:::||||::.:|||
 601  ..............ProSerGlyProAlaGlyGluArgGlyGluGlnGlyA 613

 658  GCTGCCTCACTGTTGTTGGCCTGGCCCGGGTCAGGCTACAGAGTTTGT 609
      ||||||||||      :::|||||||  |||||::.:||| |||:::|||
 613  la................ProGlyPro..SerGlyPheGlnGlyLeuP 623

 608  GACGGCGGCCCATGCTGAGCCCACCAGGAGGACGCGGGAGGTGGGCGGCGGA 559
      |||       :::|||||||  ||| |||||||||||||:::
 623  roGlyProProGlyProProGlyLysGlyIleGlyLysGlnGlyAspGlnGly 639

 558  GGACCTTCCGTGGGAGTGGCAGGGTCCCCC.......AAAGCACGCTT 518
      |||              :::|||||::.:|||       |||
 640  IlePro......GlyGluAlaGlyAlaProGlyLeuValGlyProArgGl 654

 517  GGAGTTGCCATTTTGGTCCGAGTCCAGGCGCCTGTGGAGAGCCTCTCT 468
      ||||||:::|||            :::||::||||:::
 654  yGluArgGlyPheProGlyGluArgGlySerProGlyAlaGlnGlyLeuG 671

 467  GAAGTCCTTAGATAGACAACAGTGGCAGAGAGCTGCATCTAGATGG 418
      :::.:|||||                         |||||
 671  lnGlyPro................................ArgGlyLeuPro.. 677

 417  CTGTGAGAGGCAGGAGGGTTCCTGGG..............GGCTCAGC 383
      |||            |||:::            ||||||  |||::.:||
 678  ..............GlyThrProGlyThrAspGlyProLysGlyAlaAl 689

 382  GGGCTCTGAGCTGCACACCTGCAGGGCCCCAGGCGACCCTCCCTAGTAG 333
      |||||            ||||||  |||||||:::    |||||
 689  aGlyProAsp..Gly.ProProGlyAlaGln...GlyPro.......... 700

 332  TAGCATCCACAGCTACAGGCAAGTGAAGCAGGAGGGGCAGGAGCTGACC 283
      :::||||            |||:::||| ||||
 701  ......ProGlyLeuGlnGlyMetProGlyGluArg.GlyAlaAlaGlyI 715

 282  TTACTTGTCTGAAGGGCTCGAAGCCA...................... 257
      ||||||            |||||||:::
 715  leAlaGlyProLysGlyAspArgGlyAspValGlyGluLysGlyProGlu 731

 256  ...............................................GCACC 252
      ||||||
 732  GlyAlaProGlyLysAspGlyArgProGlyValArgGlyLeuThrGlyProIleGlyPr 748
```

```
251 TCCGGGCATGGCCTGTGAACCAAGGAGCAGTCTGTGGAGCCCCACCA.  203
    ||||||  ||  ||::::::  |||:|::::        ||||||
748 oProGlyProAlaGlyAlaAsnGlyGluLysGlyAlaGlyProProG  765

202 .........CCCACCCTCTACAGAGGGCTCATAGAAAAATTGCAGTG  164
             |||   ||||||                 :::  :::
765 lyProSerGlySerThrGlyAlaArgGlyAlaProGlyGluProGlyGlu  781

163 TTCTCCCCAGCTTCTCCCTCTATCGACTTTAACAGAGAGTAGATTGCCTTGC  114
    ::::||| :::||||::                            :::|
782 ThrGlyProProGlyProAla..................GlyPheAl  791

113 TAGGCCCAGAACCAGCTCAGACTCCCAACTTTGACAAGAAGTGATCAGG  64
    |  |||   :::::   :::|||   ::::::::||||||
791 aGlyProProGlyAlaAspGlyGlnProGly...AlaLysGlyAspGlnG  807

63 CTGAATCTTCAACCCATAGCAGAGCAAGAGAGCCCCC  29
   ::|||::::::    :::  |||   ||||||
807 lyGluAlaGlyGlnLysGlyAspAlaGlyAlaPro  818
```